DOCTOR OF PHILOSOPHY

Structure of T6SS components and assessment of potential Gram-ve drug targets

Rao, Vincenzo A.

Award date:
2012

DOCTOR OF PHILOSOPHY

Structure of T6SS components and assessment of potential Gram-ve drug targets

Vincenzo A. Rao

2012

University of Dundee

A thesis submitted for the degree of Doctor of Philosophy

Structure of T6SS components and assessment of potential Gram-ve drug targets

Vincenzo A. Rao

Supervisor:

Professor William N. Hunter

The Wellcome Trust Biocentre

University of Dundee

Scotland

August 2012

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. Bill Hunter for the opportunity to work in his lab and for his support and encouragement throughout my PhD. A thanks goes to all the WNH lab members, past and present, for making my time in the lab an enjoyable and unforgettable experience and for their advice and assistance throughout the years. In particular, I would like to thank Thomas Eadsforth for his support, guidance and friendship. A thanks also goes to both Sarah Coulthurst and Grant English for their collaboration and discussions on all things related to the T6SS.

I would like to thank my family for their support while writing this thesis and the understanding they have all shown me. In particular, I would like to thank my brother Francesco for his encouragement and discussions.

Finally, a special thanks goes to my fiancée Lynsey for her support and for always being there for me. Thank you.

Declaration

I hereby declare that the following thesis is based on the result of investigations conducted by myself, and that this thesis is of my own composition. This thesis has not, in whole or in part, been previously presented for a higher degree. Work other than my own is clearly indicated in the text by reference to the relevant researchers or publications.

Vincenzo A. Rao

The work presented in this thesis is the work of the candidate V. A. Rao. Conditions of the relevant Ordnance and Regulations have been fulfilled.

Prof. W.N. Hunter

SUMMARY

The structure and potential function of two components of the type VI secretion system in *Serratia marcescens* have been investigated by X-ray crystallography. Chapters 1 and 2 describe the background of the type VI secretion system and experimental methods respectively. In chapter 3, the solution of the structure of *S. marcescens* Lip is presented. This is an essential outer membrane lipoprotein that is conserved among bacterial species. It is part of a membrane-spanning sub-assembly that is thought to anchor the remaining components to the cell envelope. The crystal structure of *Sm*Lip was determined using single-wavelength anomalous dispersion methods targeting iodides that were present in the crystallisation solution. The structure was refined to 1.92 Å and displays a fold similar to that of transthyretin; a protein that binds the hormone thyroxine. Despite the sequence identity between the two proteins being low, *Sm*Lip can be considered a new member of the transthyretin-like protein family. A role is suggested for loop 1 being involved in protein-protein interactions with other components of the type VI secretion system and therefore a potential drug target.

In chapter 4, the structure of *S. marcescens* Rap1b is described. This protein is thought to be part of a set of periplasmic resistance proteins that protect *S. marcescens* from succumbing to the effects of its own secreted toxins. The crystal structure of Rap1b was determined using single-wavelength anomalous dispersion methods targeting both iodides from a halide soak and endogenous sulfurs. The structure was refined to 1.88 Å and displays a helical rich fold not previously characterised. This structure is described and is suggested to represent this family of resistance proteins.

In addition to determining the structure of components of the type VI secretion system, the assessment of two drug targets have been investigated by biochemical techniques.

The need for new antimicrobials and the importance of target selection in the drug discovery process is described in chapter 5. PBP3 from both *Pseudomonas aeruginosa* and *Burkholderia pseudomallei* is the focus of chapter 6. This protein is essential in bacteria as it catalyses the cross-linking of peptidoglycan. Peptidoglycan provides bacteria with structural rigidity and strength to withstand pressures that would otherwise rupture the cell. Sequence identity between the two homologues is approximately 40 %, with the catalytic residues conserved. The production of soluble protein and the results of screening for novel chemical fragments using several techniques are presented.

The characterisation of *P. aeruginosa* IMPDH, an essential protein that is present in nearly all living organisms is described in chapter 7. This protein controls the pool of guanine nucleotides within cells. Fragments were identified using bio-layer interterformetry, a label free optical analytical technique that monitors biomolecular interactions in real time. Fragments were assessed for inhibitory effects using enzyme kinetics. The crystal structure of *Pa*IMPDH was determined by molecular replacement and refined to 2.25 Å. This structure displays the catalytic domain with the active site loop fully ordered and in a conformation rarely observed.

ABBREVIATIONS

SI units are used throughout with the exception of Å representing 10^{-10} m.

AAA^+	ATPases associated with various cellular activities
AI	Auto-induction media
ASA	Accessible surface area
APBS	Adaptive poisson-boltzmann solver
ATCC	American type culture collection
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate
BAP	Biotin acceptor peptide
BirA	Biotin-protein ligase
Bp	Base pair
BLI	Bio-layer interferometry
CBS	Cystathionine β-synthase
CCS	Clinical candidate selection
DDU	Drug discovery unit
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
DNTP	Deoxyribonucleotide triphosphate
DPI	Diffraction precision index
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
EAEC	Enteroaggregative *E. coli*
EDTA	Ethylenediamine tetraacetic acid

ESRF	European synchrotron radiation facility
GlcNAc	*N*-acetylglucosamine
GMP	Guanine 5'-monophosphate
GPRT	Guanine phosphoribosyl transferase
dGTP/GTP	Deoxyguanosine/guanosine triphosphate
HAI	Hospital-acquired infection
Hcp	Hemolysin co-regulated protein
HMM	High molecular mass
HSI	Hcp-secretion island
IM	Inner membrane
IMPDH	Inosine 5'-monophosphate dehydrogenase
IMP	Inosine 5'-monophosphate
IPTG	Isopropyl β-D-thiogalactoside
ITC	Isothermal titration calorimetry
kDa	Kilodalton
K_d	Dissociation constant
LB	Lysogeny broth
LMM	Low molecular mass
LSI	Lead series identified
MAD	Multi-wavelength anomalous dispersion
MALDI-TOF	Matrix-assisted, laser deorption/ionisation time-of-flight
MES	2-(*N*-morpholino)ethanesulfonic acid
MIR	Multiple isomorphous replacement
MMF	Mycophenolate mofetil
MPA	Mycophenolic acid
MPD	2-methyl-2,4-pentanediol

mesoA_2pm	Mesodiaminopimelic acid
MurNAc	*N*-acetylmuramic acid
NAD^+/H	Nicotinamide adenine dinucleotide
NCS	Non-crystallographic symmetry
NK	Nucleoside kinase
NMR	Nuclear magnetic resonance
OM	Outer membrane
OD	Optical density
OPPF	Oxford protein production facility
PAGE	Polyacrylamide gel electrophoresis
PBP	Penicillin-binding protein
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PEP	Phosphoenolpyruvate
pI	Isoelectric point
Rap	Resistance associated protein
Rcf	Relative centrifugal force
RMSD	Root mean square deviation
RNA	Ribonucleic acid
SAD	Single-wavelength anomalous dispersion
SDS	Sodium dodecyl sulfate
Sec	General secretory pathway
ssDNA	Single-stranded DNA
Ssp1-2	Small secreted protein 1-2
SPR	Surface plasmon resonance

Tat	Twin-arginine translocation pathway
TEV	Tobacco-etch virus
TIM	Triose-phosphate isomerase
TLS	Translation/libration/screw
T_m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
Tse1-3	Type VI secretion exported proteins 1-3
Tsi1-3	Type VI secretion immunity proteins 1-3
T1-6SS	Type I-VI secretion system
TssA-M	Type six secretion proteins A-M
Tag	Tss associated genes
UDP	Uridine diphosphate
UP	Undecaprenyl phosphate
UTI	Urinary tract infection
VgrG	Valine glycine repeat protein G
VHS	Validated hit series
XMP	Xanthosine 5'-monophosphate
XPRT	Xanthine phosphoribosyl transferase

PART I

STRUCTURE OF T6SS COMPONENTS

CHAPTER 1

INTRODUCTION: The type VI secretion system

1.1 Secretion in Gram-negative bacteria

Protein secretion is the transport of proteins from the cytosol to the extracellular milieu and is a basic cellular function in both Gram-positive and Gram-negative bacteria (Gerlach and Hensel, 2007). While the secretion of proteins in Gram-positive bacteria has only the single membrane barrier to traverse, Gram-negative bacteria must overcome the inherent difficulty of a second membrane (**Figure 1.1**). These proteins must traverse the inner membrane (IM), the periplasm and then the outer membrane (OM) of the cell envelope before secretion is achieved. In some instances, the intended target of the secreted protein is a target cell, adding an additional membrane barrier to overcome (Fronzes *et al*., 2009; Mueller *et al*., 2008). In general secreted proteins have a multitude of functions, including the biogenesis of organelles such as pili and flagella, intercellular communication, nutrient acquisition, virulence, protein adhesins, effector proteins and the efflux of drugs (Gerlach and Hensel, 2007; Kostakioti *et al*., 2005). Protein secretion in Gram-negative bacteria is therefore associated with bacterial fitness, competition and also with virulence of pathogenic bacteria. The understanding of the mechanisms involved in protein secretion may provide opportunities for the development of new drug treatments against these pathogens. This is important as the increase in hospital-acquired infections from Gram-negative bacteria (Chopra *et al*., 2008; Peleg and Hooper, 2010) and instances of multi drug resistance become a growing concern (Arias and Murray, 2009; Souli *et al*., 2008).

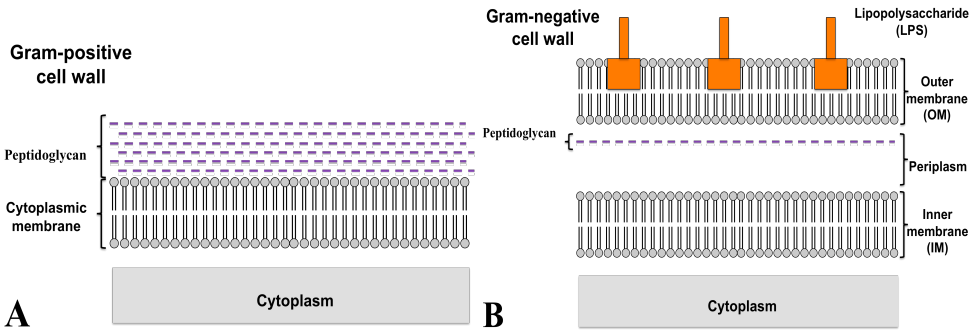


Figure 1.1 The cell wall of bacteria.
The Gram-positive bacterial cell wall (**A**) has a single membrane present, with a large peptidoglycan layer providing rigidity and structural strength. The Gram-negative bacterial cell wall (**B**) has a second membrane present, sandwiching the comparatively smaller peptidoglycan layer. Note that porins and membrane embedded proteins are excluded for clarity. Lipid A is shown in orange.

1.1.1 An overview of protein secretion systems

Macromolecular complexes known as secretion systems have evolved in Gram-negative bacteria to facilitate protein secretion (Holland, 2010). One step or two step mechanisms are used to secrete proteins to the exterior of the cell (**Figure 1.2;** Holland, 2010). Six secretion systems have been identified in Gram-negative bacteria (referred to as T1SS-T6SS), comprising of 3 to more than 20 subunits. The sixth secretion system is the subject of the studies presented in this thesis. All of the secretion systems have a different composition and arrangement of structural components, therefore utilising distinct mechanisms (Gerlach and Hensel, 2007; Holland, 2010).

The different secretion systems can broadly be divided into two categories; those independent of the general secretory pathway (Sec) and twin-arginine translocation (Tat) pathway, and those that utilise the Sec or Tat pathway (**Figure 1.2;** Gerlach and Hensel, 2007). The T1SS secretes proteins directly from the cytoplasm to the exterior of the cell in a single step, with typical substrates being high-molecular weight toxins such as HlyA (Holland *et al*., 2005, Thanabalu *et al*., 1998). Substrates of the T2SS are targeted to the Sec or Tat pathway via a signal peptide sequence in order to traverse the inner membrane. Fully folded protein is then secreted across the periplasm and outer

membrane to the extracellular environment using the T2SS (Johnson *et al*., 2006). The T3SS forms both a membrane and periplasm-spanning complex, with an external needle like structure protruding from the cell surface. This system secretes bacterial virulence factors directly into host cells (Mueller *et al*., 2008). The T4SS exports both proteins and single-stranded DNA (ssDNA) in a manner similar to bacterial conjugation (Fronzes *et al*., 2009; Kostakioti *et al*., 2005). The T5SS, which includes the autotransporter system and two partner secretion pathways, transports substrates in two steps, utilising the Sec pathway initially. Once in the periplasm, proteins insert themselves into the outer membrane via their C-terminal domain, promoting translocation of the N-terminal domain across the outer membrane (Henderson *et al*., 2004).

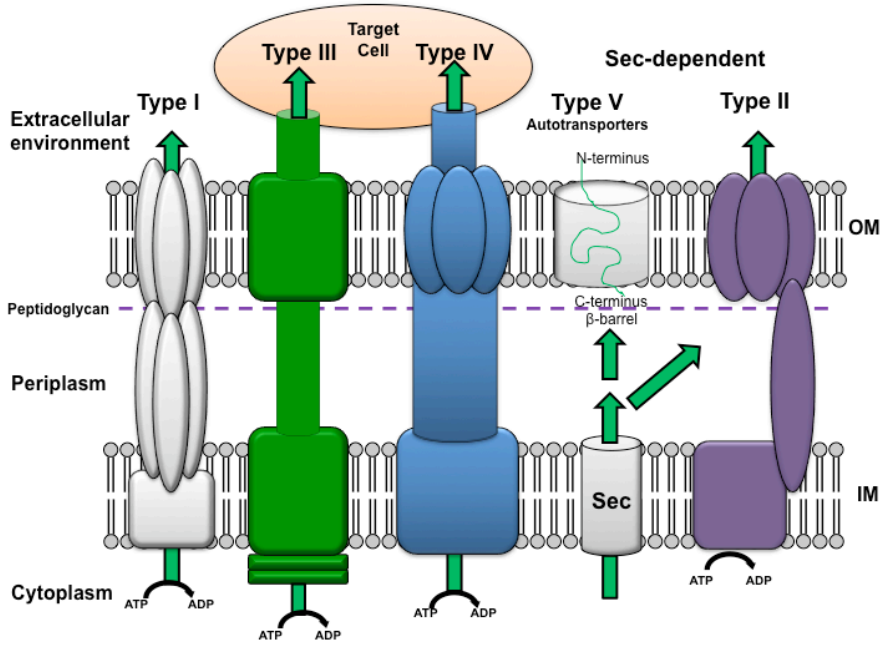


Figure 1.2 Secretion systems.
Secretion systems are macromolecular complexes consisting of several subunits varying from a simple membrane-spanning channel (T1SS) to more complex systems that act as molecular syringes (T4SS). Figure adapted from (Desvaux *et al*., 2009).

1.2 The type VI secretion system

The type VI secretion system (T6SS) is the most recently identified of the Gram-negative secretion systems and the least characterised. First identified in 2006 by its role in preventing predation of *Vibrio cholerae* by amoebae (Pukatzki *et al*., 2006), the T6SS is now recognised to be involved in a number of bacterial phenotypes (Jani and Cotter, 2010). These include quorum sensing in *Vibrio anguillarum* (Weber *et al*., 2009), virulence of *Burkholderia pseudomallei* towards mice (Chen *et al*., 2011), biofilm formation by enteroaggregative *Escherichia coli* (EAEC) (Aschtgen *et al*., 2008), and antibacterial activity of *Pseudomonas aeruginosa* and *Serratia marcescens* (Murdoch *et al*., 2011; Russell *et al*., 2011). The gene cluster encoding a T6SS is found in the genome of approximately 25 % of Gram-negative bacteria (Bingle *et al*., 2008). Typically these clusters contain more than 20 genes, with the composition and organisation of these genes varying between bacterial species. Several species have a number of gene clusters corresponding to multiple T6SSs (Bingle *et al*., 2008; Boyer *et al*., 2009). For example, the genome of *P. aeruginosa* has three T6SS clusters (HSI-1 – HSI-3; Mougous *et al*., 2006) that are thought to have been acquired by horizontal gene transfer (Bingle *et al*., 2008). These clusters are not co-regulated, suggesting they each carry out different functions (Bingle *et al*., 2008). While the HSI-1 T6SS is known to secrete bacteriolytic effector proteins for a competitive advantage (Russell *et al*., 2011), the HSI-2 T6SS is reported to be involved in interactions with eukaryotic cells (Sana *et. al*., 2012). It is currently unclear what the function of the HSI-3 T6SS cluster is.

Approximately 13 conserved genes have been identified as the minimum required for a functional T6SS, representing the core components of the secretion apparatus (**Table 1.1**; Boyer *et al*., 2009). Although the function of many of these components has yet to be determined, the genes have been validated as essential in several bacteria (Zheng *et al*., 2011; Zheng and Leung, 2007). In an attempt to unify the

nomenclature of these conserved genes, they have been named Tss (type six secretion) A – M (**Table 1.1**), while auxiliary components of the T6SS are designated Tag (Tss associated genes; Shalom *et al*., 2007). The common name of characterised components, where appropriate, will be used herein.

Table 1.1 Nomenclature of T6SS components.

Unified name	Common name	Information	Homologue	Associated reference
TssA		Putative cytosolic protein		
TssB	VipA	Forms a large cogwheel like tube surrounding Hcp	Bacteriophage sheath	(Bönemann *et al*., 2009)
TssC	VipB			
TssD	Hcp	Forms hexameric rings, putative pilus, always present in extracellular media	Bacteriophage tube gp19	(Bingle *et al*., 2008; Mougous *et al*., 2006)
TssE		Putative soluble protein		
TssF		Putative soluble protein		
TssG		Putative soluble protein		
TssH	ClpV	AAA^+ ATPase required to disassemble the VipA/B complex	Hsp100	(Pietrosiuk *et al*., 2011)
TssI	VgrG	Cell puncturing device	Bacteriophage spike gp27/gp5	(Leiman *et al*., 2009)
TssJ	SciN/Lip	Outer membrane lipoprotein		(Felisberto-Rodrigues *et al*., 2011; Rao *et al*., 2011)
TssK		Putative cytosolic protein		
TssL	IcmH	Inner membrane protein	T4bSS IcmH	(Durand *et al*., 2012)
TssM	IcmF	Inner membrane protein with ATPase activity	T4bSS IcmF	(Ma *et al*., 2012)

Many of the core components of the T6SS can be divided into two groups. The first is components that share structural similarities to bacteriophage tail tube, sheath or spike proteins (Leiman *et al*., 2009; Pukatzki *et al*., 2007). These proteins are thought to form a sub-assembly that resembles an inverted bacteriophage, forming the injection apparatus of the T6SS (**Figure 1.3**; Basler *et al*., 2012; Bönemann *et al*., 2010; Cascales, 2008). The second group is components that are embedded, or associated, with the inner or outer membrane. These proteins form a membrane-spanning sub-

assembly that is thought to anchor the bacteriophage-like injection apparatus to the bacterial cell envelope (Aschtgen *et al*., 2010b; Durand *et al*., 2012; Felisberto-Rodrigues *et al*., 2011).

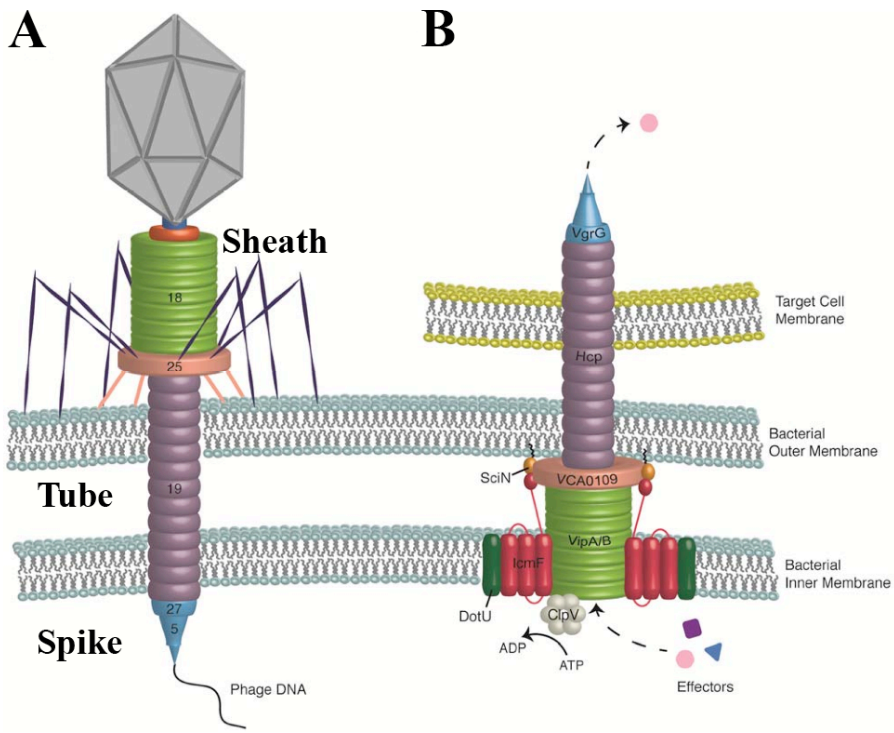


Figure 1.3 Comparison with bacteriophage.
A) Schematic representation of the T4 bacteriophage tail components and their arrangement. **B**) The proposed model of the T6SS. Equivalent components are coloured the same. Figure edited from (Records, 2011).

1.2.1 Assembly and function: the bacteriophage-like injection apparatus

A hallmark of the T6SS is the presence of the protein hemolysin co-regulated protein (Hcp) in culture supernatant (Bingle *et al*., 2008; Cascales, 2008). Hcp is a conserved and essential component of the T6SS and is used as an indicator of a fully functional system (Zheng and Leung, 2007). The presence of Hcp is mutually dependent upon the protein valine glycine repeat protein G (VgrG); deletion of either Hcp or VgrG results in the absence of the other protein in culture supernatant (Pukatzki *et al*., 2007; Zheng and Leung, 2007).

The crystal structure of Hcp reveals six molecules assembled to form a hexameric ring approximately 80 Å wide with an internal diameter of approximately 40

Å, large enough to accommodate a small folded or unfolded protein (PDB 1Y12; **Figure 1.4A**, Mougous *et al*., 2006). Analysis of Hcp by electron microscopy reveals the Hcp hexamers assemble to form a tubular structure of approximately 100 nm in length (**Figure 1.4A**; Ballister *et al*., 2008). This nanotube spans the inner membrane, periplasm and outer membrane, possibly functioning as a conduit to translocate substrate from the cytosol to the extracellular milieu. The tertiary structure of Hcp displays structural similarity to that of bacteriophage tail tube protein (PDB 2K4Q; Leiman *et al*., 2009; Pell *et al*., 2009).

The structure of the N-terminus of VgrG reveals a trimer (PDB 2P5Z; **Figure 1.4B**), with sequence and structural comparisons indicating homology with bacteriophage tail spike proteins, in particular the proteins gp27 and gp5 (PDB 1K28; **Figure 1.4C**, Kanamaru *et al*., 1994; Leiman *et al*., 2009; Pukatzki *et al*., 2007). This has led to the suggestion that the T6SS functions similar to an inverted bacteriophage (**Figure 1.3**), with VgrG functioning as a cell-puncturing device located on the tip of the Hcp nanotube (**Figure 1.4D**). The presence of Hcp and VgrG in culture supernatant may therefore be a result of the Hcp nanotube extending and pushing VgrG out of the cell to the external environment, allowing VgrG to puncture target cells. As VgrG lacks an internal channel, the delivery of substrate would be achieved following disassembly of VgrG from the Hcp nanotube.

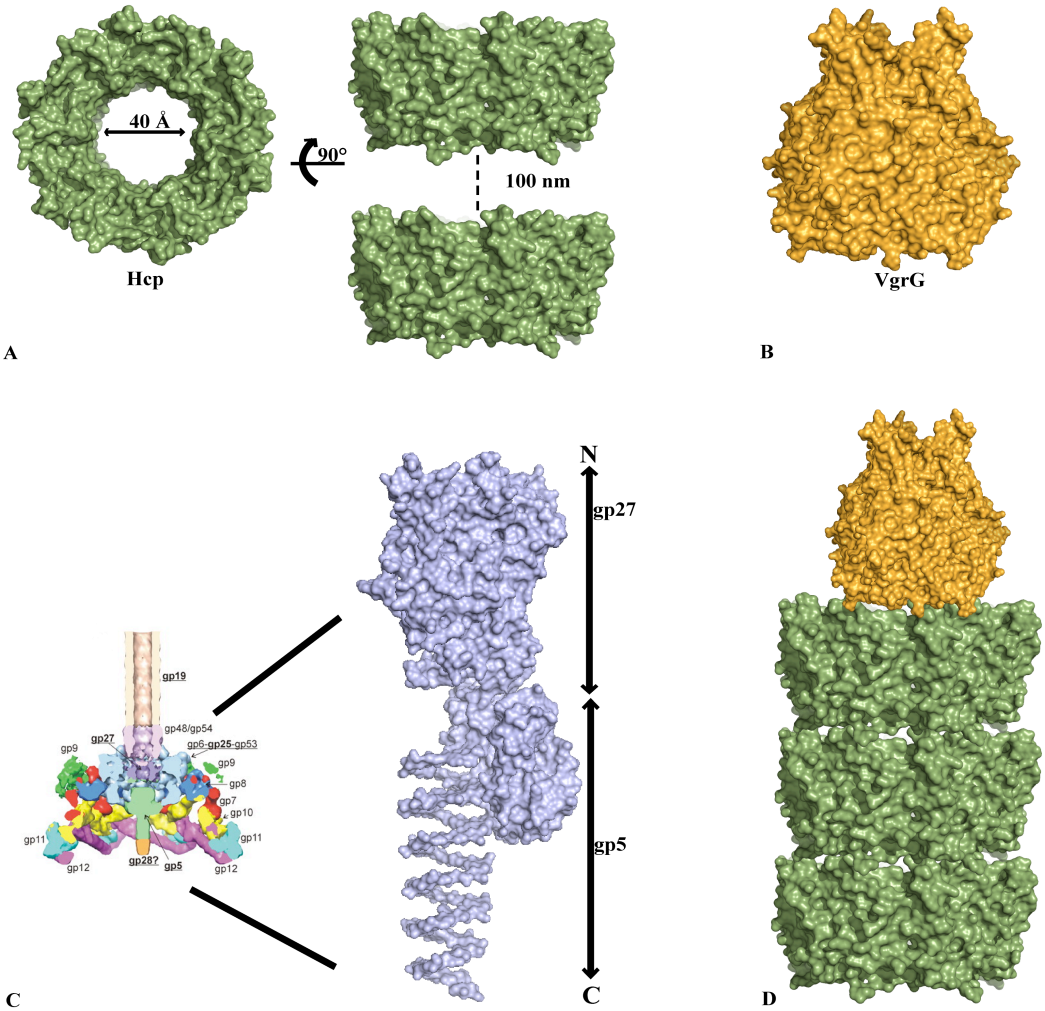


Figure 1.4 The T6SS injection apparatus.
A) The structure of Hcp forms a hexameric ring (PDB 1Y12; Mougous *et al*., 2006). These rings are able to assemble into long nanotubes that span the inner and outer membranes (Ballister *et al*., 2008). **B**) The structure of VgrG forms a trimer (PDB 2P5Z; Leiman *et al*., 2009) with similarities to the bacteriophage tail spike proteins gp27 and gp5 (PDB 1K28; **C,** Kanamaru *et al*., 1994). **D**) The Hcp nanotube harbouring VgrG on the tip to form a cell-puncturing device.

1.2.2 Assembly and function: the membrane complex

Three conserved proteins in the T6SS are membrane-bound and form a large trans-membrane complex, anchoring the bacteriophage-like injection apparatus to the cell envelop (Aschtgen *et al*., 2010b; Cascales and Cambillau, 2012; Durand *et al*., 2012).

The proteins IcmH and IcmF are localised to the inner membrane and have similarities to components within the T4SS (Bingle *et al*., 2008; Durand *et al*., 2012). IcmH is inserted into the inner membrane via a single transmembrane helix at the C-terminus (Aschtgen *et al*., 2011). In most cases, this protein has a C-terminal

periplasmic domain that encompasses a peptidoglycan-binding motif. Instances where this periplasmic domain is absent, an auxiliary protein, TagL, compensates by possessing this binding motif and interacting with IcmH (Aschtgen *et al*., 2010a; Aschtgen *et al*., 2010b). A major portion of IcmH is located in the cytoplasm (Van Rheenen *et al*., 2004) and the crystal structure of this domain has recently been solved (PDB 3U66; **Figure 1.5A**, Durand *et al*., 2012). IcmH has eight α-helices comprising two three-helix bundles that are connected via two short α-helices. The function of this domain is unclear, it is possible this domain may act to aid in the recruitment of substrates to the T6SS (Durand *et al*., 2012), however there is no evidence for this.

The protein IcmF has several cytoplasmic residues at the N-terminus, followed by two transmembrane helices separated by a short loop and a cytoplasmic domain of approximately 330 residues (Cascales and Cambillau, 2012). This domain has Walker A and B motifs and is involved in energising the T6SS, possibly with the ATPase ClpV (PDB 3ZRI; **Figure 1.5B**, Ma *et al*., 2012; Pietrosiuk *et al*., 2011). A third transmembrane helix follows the cytoplasmic domain and a large periplasmic domain of approximately 740 residues comprises the rest of the protein. IcmF has been shown to interact with IcmH via their respective cytoplasmic domains (Ma *et al*., 2009; Zheng and Leung, 2007). In addition, IcmF has recently been shown via surface plasmon resonance (SPR) to interact with Lip, a lipoprotein located at the outer membrane (Felisberto-Rodrigues *et al*., 2011). Therefore IcmF links both the inner and outer membranes, forming a large trans-membrane complex (**Figure 1.5C**).

Our understanding of how T6SSs function has improved recently and has led to a proposed mechanism (Basler *et al*., 2012). It is thought that the conserved proteins VipA/VipB (**Table 1.1**, **Figure 1.3**) are homologous to the bacteriophage tail sheath protein gp18 (Aksyuk *et al*., 2009; Basler *et al*., 2012; Bönemann *et al*., 2010; Leiman *et al*., 2009). VipA/VipB polymerise to form a tubule with an inner diameter of

approximately 100 Å, large enough to encompass the nanotube of Hcp (approximately 80 Å; Bönemann *et al*., 2009; Mougous *et al*., 2006). In bacteriophage, the tail sheath protein contracts to provide the energy required to push the tail spike through the outer membrane of target cells (**Figure 1.3**; Aksyuk *et al*., 2009). Similarly, the proteins VipA/VipB are able to assemble and contract to energise the expulsion of the Hcp nanotube harbouring VgrG, allowing the injection of substrates into target cells (Basler *et al*., 2012). This contractile bacteriophage tail-like structure has been shown to extend to the approximate length of the cell, before contraction to about 50 % of the extended length (Basler *et al*., 2012). The VipA/VipB sheath is then disassembled by ClpV (**Figure 1.5B**) via an ATP-dependent remodelling process (Bönemann *et al*., 2009; Pietrosiuk *et al*., 2011). ClpV, a member of the Hsp100/Clp family of AAA^+ (ATPases associated with various cellular activities) proteins, forms a hexameric ring through which the VipA/VipB complex is threaded through and dissassembled upon ATP hydrolysis (Bönemann *et al*., 2009). The N-terminual domain of ClpV confers substrate specificity and mediates the binding with VipA/VipB subunits (Pietrosiuk *et al*., 2011).

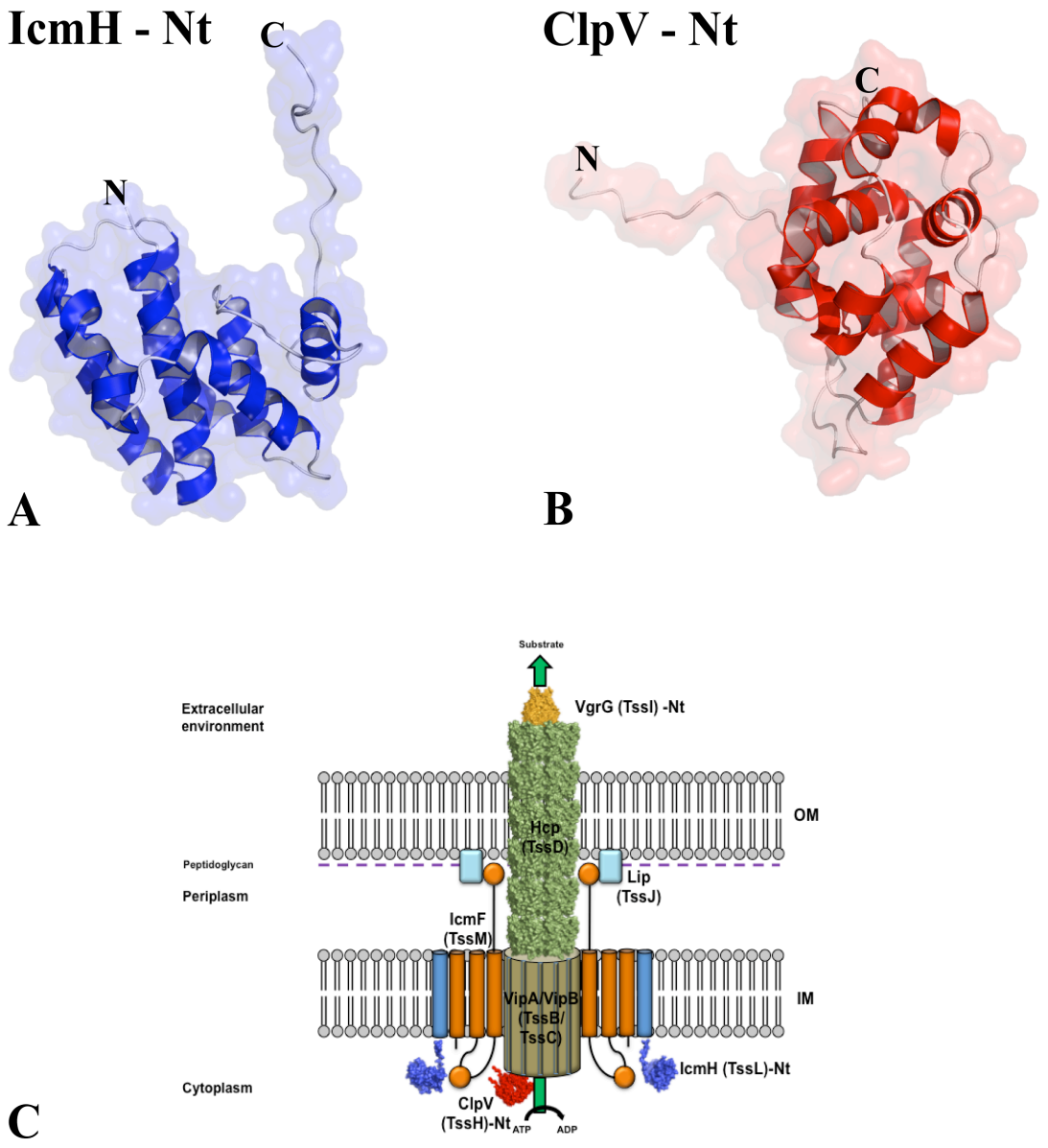


Figure 1.5 The T6SS.
A) The structure of the N-terminal cytoplasmic domain of IcmH (PDB 3U66; Durand *et al*., 2012). **B**) The N-terminal domain of the AAA^{+} ATPase ClpV. This domain is responsible for substrate specificity (PDB 3ZRI; Pietrosiuk *et al*., 2011). **C)** A schematic overview of the T6SS showing the postulated arrangement and interactions of several conserved components. Figure adapted from (Bönemann *et al*., 2010; Cascales and Cambillau, 2012).

1.3 *S. marcescens* Db10 T6SS

A functional T6SS has been identified in *S. marcescens* strain Db10 (**Figure 1.6**), where this system is constitutively expressed (Murdoch *et al*., 2011). This system exhibits antibacterial activity against several bacterial species, including *E. coli*, *Enterobacter cloacae* and *P. fluorescens* (Murdoch *et al*., 2011). Significantly, this system is able to target and kill *S. marcescens* ATCC 274, a closely related strain with a similar T6SS

(Murdoch *et al*., 2011). This suggests the Db10 strain secretes effector proteins to which the ATCC 274 strain lacks resistance against. Self-resistance proteins must therefore be present in the Db10 strain, allowing the bacteria to survive and out-compete other bacteria. This would be similar to the antibacterial *P. aeruginosa* HSI-1 T6SS, where the immunity proteins Tsi1-3 (Type VI secretion immunity) negate the effects of the secreted toxins Tse1-3 (Type VI secretion exported; Hood *et al*., 2010; Russell *et al*., 2011). These proteins are restricted to *P. aeruginosa*, allowing the targeting and killing of closely related species to give a competitive advantage (Russell *et al*., 2011). Potential candidates for similar proteins in the T6SS of *S. marcescens* have been identified (**Figure 1.6**; Murdoch *et al*., 2011).

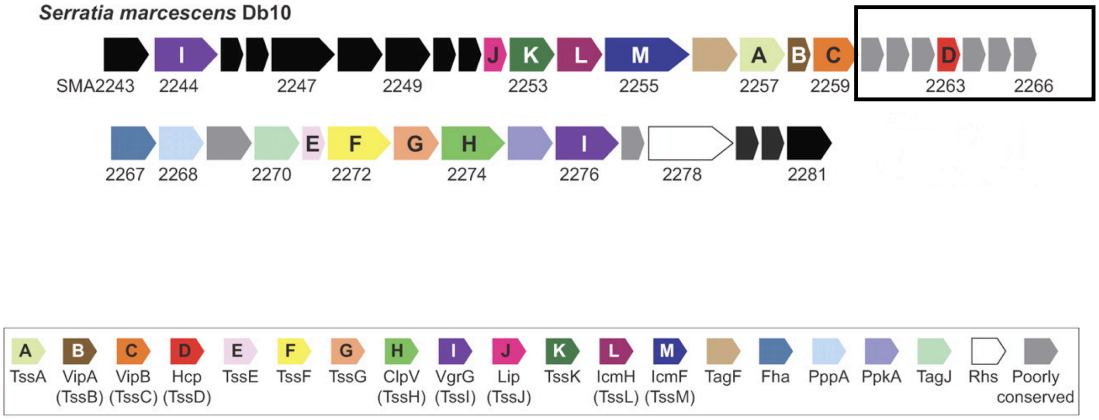


Figure 1.6 Schematic representation of the T6SS genes in *S. marcescens*.
Genes are colour coded and labelled (top) according to the gene name (*SMA2243 – SMA2281*) and letters using the Tss unified nomenclature (Boyer *et al*., 2009). A key (bottom) shows the colour, Tss unified name and the common name of conserved components of the T6SS. Potential candidates for secreted effector / self-resistance proteins are boxed. Figure edited from (Murdoch *et al*., 2011).

1.3.1 Secreted toxins and self-resistance proteins

A number of genes with previously unknown function have been identified in the middle of the T6SS operon in *S. marcescens* (**Figure 1.6**; Murdoch *et al*., 2011). These genes (*SMA2260 – 2262* and *SMA2264 – 2266*; **Figure 1.7**) encode small proteins with homologues in several other T6SSs, including those of *Cronobacter sakazakii* and *E. cloacae* (Murdoch *et al*., 2011). These small proteins exhibit low sequence identity to

one another, yet they can be classified into two distinct groups based on their isoelectic point (pI), secondary structure prediction and the presence of a signal peptide sequence (**Figure 1.7**).

The basic proteins, *Sm*2261 and *Sm*2264, are predicted to be mainly helical with several strands and lack a signal peptide sequence. These proteins represent secreted toxins (English *et al.*, 2012) and have been termed Ssp1 and Ssp2 (Small secreted protein; **Figure 1.7**) respectively. These toxins contribute to the antibacterial activity associated with this secretion system (English *et al.*, 2012; Murdoch *et al*., 2011) and are analogous to the Tse1 and Tse3 toxins in the *P. aeruginosa* HSI-1 T6SS (Hood *et al*., 2010; Russell *et al*., 2011). Tse1 and Tse3 are peptidoglycan hydrolases delivered to the periplasm of a target cell (Russell *et al*., 2011). Ssp1 and Ssp2 exhibit no sequence similarity to other peptidoglycan hydrolases, yet both have been shown to be periplasmic-acting toxins with distinct functions (English *et al*., 2012).

The remaining four acidic proteins (*Sm*2260, *Sm*2262, *Sm*2265 and *Sm*2266) represent cognate self-resistance proteins (English *et al.,* 2012), providing protection against Ssp1 and Ssp2. These helical rich proteins have therefore been termed Rap (Resistance associated protein; **Figure 1.7**) and are located to the periplasm in order to bind and prevent the periplasmic Ssp proteins from harming the producing bacteria (English *et al.,* 2012). This interaction between Ssp and Rap proteins has been shown to be in the nanomolar range and displays high specificity, with Ssp1 and Ssp2 interacting with Rap1a and Rap2a respectively (English *et al.,* 2012). The function of the remaining Rap proteins, Rap1b and Rap2b, is thought to provide protection to either unidentified additional toxins, or against bacteria secreting similar toxins (English *et al.,* 2012).

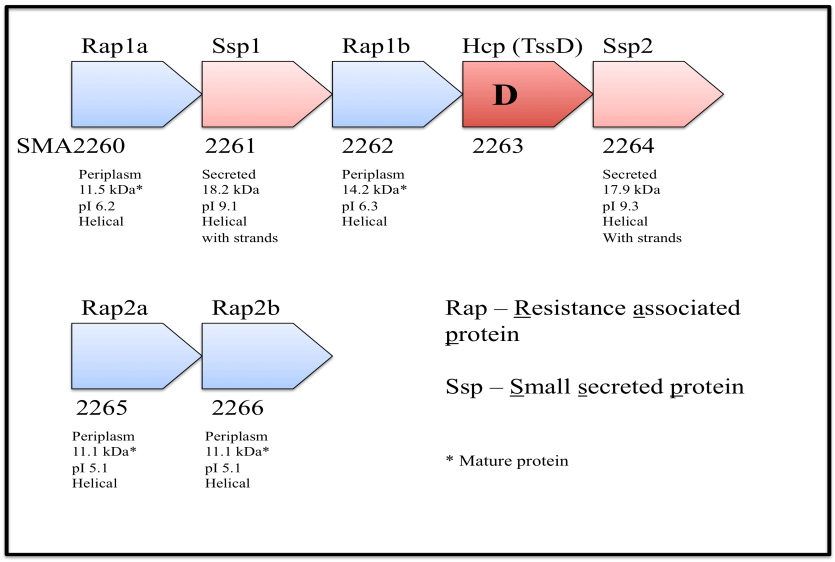


Figure 1.7 Schematic representation of the genes encoding several small proteins.
These small proteins are analogous to the Tse secreted toxins and Tsi resistance proteins in the HSI-1 T6SS of *P. aeruginosa* (Hood *et al*., 2010; Russell *et al*., 2011). Figure adapted from (English *et al*., 2012).

1.3.2 *Serratia marcescens*

S. marcescens is a Gram-negative pathogen first described in the early 19th century due to its characteristic red pigment (Bizio, 1823). The bacterium targets immunocompromised individuals, causing a broad range of diseases including urinary tract infections (UTIs), pneumonia and meningitis (Hejazi and Falkiner, 1997). Infections are difficult to treat as *S. marcescens* is naturally resistant to most β-lactam antibiotics (Ball *et al*., 1977; Mahlen, 2011; Sleigh, 1983; Wheat *et al*., 1951). A survey carried out in 2008 revealed the *Serratia* species represents 2.0 % of all hospital-acquired bloodstream infections and 2.8 % of all instances of hospital-acquired pneumonia in Europe (European Centre for Disease Prevention and Control; Sprenger, 2010). With *S. marcescens* recognised as the most common *Serratia* species isolated from human infections (Laupland *et al*., 2008), this pathogen is a significant cause of hospital-acquired infections (Hejazi and Falkiner, 1997). It is also a tractable model

organism to study the structure-function relationships in the T6SS (Murdoch *et al*., 2011).

1.4 Aims

Recent progress has been made in the structural and biochemical characterisation of the components of the T6SS (Basler *et al*., 2012; Cascales and Cambillau, 2012). This has resulted in the identification of two sub-assemblies, the first resembling the contractile tail of bacteriophage and the other a membrane spanning complex anchoring the first sub-assembly to the cell envelope. Individual deletion of several conserved components has demonstrated their involvement in forming a functioning T6SS (Zheng *et al*., 2011; Zheng and Leung, 2007). The interconnected and critical nature of these components for proper assembly of the T6SS makes them potential drug targets to explore further.

At the onset of the research carried out in this thesis, two crystal structures were available of T6SS components (Leiman *et al*., 2009; Mougous *et al*., 2006). The broad aim of this project was therefore to elucidate the structures of the other conserved components from the *S. marcescens* T6SS in the pursuit of understanding their function. In particular, the work described in chapter three was directed towards determining the structure of Lip (TssJ), an essential component of the T6SS in *S. marcescens* (Murdoch *et al*., 2011). The successful structure determination would allow us to infer its function through structural analysis and comparison with structures deposited in the Protein Data Bank (Berman *et al*., 2002). This would pave the way for further structural and biochemical studies to identify possible binding partners and obtain complexes, leading to the development of inhibitors that can potentially disrupt the assembly of a functional system (Felise *et al*., 2008; Shahian *et al*., 2009).

In addition, a number of small proteins encoded on the T6SS operon with previously unknown function were recognised (Murdoch *et al*., 2011). Their position

amongst conserved T6SS components suggested they were associated with this system in some way and may represent potential substrates for the *S. marcescens* T6SS (Murdoch *et al*., 2011). Structures of these proteins could not be modelled using the *PHYRE* protein fold recognition server (Kelley and Sternberg, 2009). Therefore the aim of this project was to gain further understanding of these proteins and their potential role within the secretion system. By determining the crystal structures of these proteins, comparisons with the recently characterised substrates in other bacterial species (Li *et al*., 2012; Zou *et al*., 2012) would be made. This is the focus of the research described in chapter four on the protein Rap1b.

CHAPTER 2

MATERIALS AND METHODS

2.1 General materials

2.1.1 Reagents

All chemicals were of analytical grade and purchased from Sigma-Aldrich, BDH Chemicals, Merck or Fluka.

Tobacco-Etch Virus (TEV) was made in-house by Kerri Barrack (University of Dundee) and provided ready to use at a concentration of 3 mg mL^{-1}. All restriction enzymes were purchased from New England Biosciences (NEB). NuPAGE 10 % Bis-Tris precast gels, MES SDS Running Buffer (20x) and SeeBlue Plus2 Prestained Standard (1x) were purchased from Invitrogen. InstantBlue gel stain was purchased from Expedeon.

2.1.2 Bacteria

S. marcescens strain Db10 genomic DNA was provided by Dr. Sarah J. Coulthurst (University of Dundee). Genomic sequence data were obtained from the Sanger Institute (www.sanger.ac.uk). The following cells were purchased from the respective companies; TOP 10 competent cells (Invitrogen), XL1 blue sub-cloning cells and *E. coli* BL21 (DE3) pLysS cells (Stratagene), *E. coli* BL21 (DE3) Rosetta-gami and *E. coli* BL21 (DE3) Gold cells (Novagen).

2.1.3 Growth media

Auto-induction (AI) media (Studier, 2005), lysogeny broth (LB) media (Bertani, 1951) and LB-agar plates were prepared, autoclaved and supplied by the media kitchen in the College of Life Sciences, University of Dundee.

2.2 General methods

Standard molecular biology techniques were used, details of which can be obtained from Molecular Cloning: A laboratory manual (Maniatis, 1983).

2.2.1 Cloning

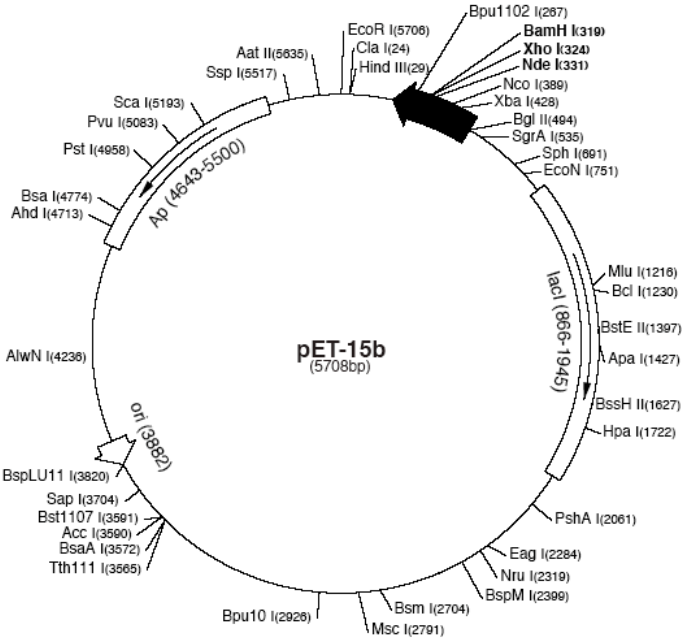
Polymerase chain reaction (PCR) was carried out to amplify the gene of interest from genomic DNA. This involves repeated heating/cooling cycles and enzymatic DNA replication using a heat stable DNA polymerase.

A fragment of the gene encoding *Sm*Lip from *S. marcescens* was cloned using *Pfu* DNA polymerase (Promega) with PCR cycling parameters given in **Table 2.1**.

Table 2.1 *Sm*Lip PCR parameters

Order of addition	Reagent	Volume added	PCR cycling conditions
1	Sterile MiliQ water	31.25 µL	95°C for 45 sec then
2	Pfu buffer 10x	5 µL	95°C for 45 sec
3	DNTPs 2 mM	5 µL	59°C for 45 sec
4	DMSO 100 %	2.5 µL	72°C for 50 sec then
5	Primers 10 pmole/µL	2.5 µL (F & R)	72°C for 10 mins
6	*Pfu* polymerase 1 unit/µL	1 µL	
7	gDNA (200 ng)	0.25 µL	
Total		**50 µL**	**36 cycles**

The gene of interest was digested out of either the PCR-BluntII-TOPO® vector (Invitrogen) or pUC57 vector, using the appropriate restriction enzymes, before ligation into a pET15b (Novagen) cloning vector (**Figure 2.1**). This vector has been modified to encode a TEV cleavage site in place of the thrombin protease cleavage site (**Figure 2.1**). Constructs were verified by DNA sequencing (DNA Sequencing Unit, University of Dundee). Plasmids were transformed into appropriate *E. coli* expression systems and the gene was expressed.




```
BglII              T7 forward primer      lac operator      XbaI                    rbs
AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA

   NcoI          His-Tag                                    NdeI  XhoIBamHI
TATACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCGAAAACCTGTATTTTCAGGGCCATATGCTCGAGGATCCGGCTGCTAACAAAGCCCG
A
     MetGlySerSerHisHisHisHisHisSerSerGlyGluAsnLeuTyrPheGln*GlyHisMetLeuGluAspProAlaAlaAsnLysAlaA
rg                                          TEV protease site

AAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
LysGluAlaGluLeuAlaAlaAlaThrAlaGluGlnEND
                          T7term primer
```

Figure. 2.1 pET15b-TEV vector.
The pET15b (Novagen) vector has been modified to include a TEV protease cleavage site, the * indicates the position where cleavage occurs. The multiple cloning sites are detailed.

2.2.2 Recombinant protein production in *E. coli*

A single colony of freshly transformed cells was used to inoculate 50 mL of LB media supplemented with carbenicillin and incubated, with aeration, at 37 °C overnight. From this starter culture, 10 mL was used to inoculate each litre of AI media with carbenicillin in a 2 litre flask. A detailed description of AI media is published (Studier, 2005). In brief, cells use the glucose in the media as the primary source of energy before switching to metabolising lactose as a secondary energy source, requiring induction of the *lac* operon. This results in the transcription and over expression of our gene of

interest. Inoculated cells were left incubating at 37 °C until an OD_{600} of 0.6 was reached. Cells were then grown overnight at 22 °C.

Cells were harvested by transferring to 1 litre plastic flasks suitable for centrifugation, before centrifuging at 3,500 rcf (Beckman J6-MC with a JA6 rotor) for 30 mins at 4 °C. The supernatant was removed and 25 mL of fresh media with carbenicillin was used to resuspend each cell pellet. The resuspension was transferred to 50 mL falcon tubes and centrifuged at 3,500 rcf (Sigma 4K15 with a 373/E rotor) for 30 mins at 4 °C. The supernatant was then discarded and the cell pellet was flash frozen in liquid nitrogen before storing at -20 °C.

2.3 Protein purification

2.3.1 Buffers

Table 2.2 gives details of the different buffers used during the purification of each protein;

Table 2.2 Buffers

Name	Buffer	Used for the purification of:
A	25 mM Tris-HCl, pH 7.5, 500 mM NaCl, 20 mM imidazole	*Sm*Lip / *Bp*PBP3 / *Pa*PBP3
B	25 mM Tris-HCl, pH 7.5, 500 mM NaCl, 500 mM imidazole	*Sm*Lip / *Pa*IMPDH / *Bp*PBP3 / *Pa*PBP3
C	25 mM Tris-HCl, pH 7.5, 250 mM NaCl	*Sm*Lip / *Pa*IMPDH / *Bp*PBP3 / *Pa*PBP3
D	25 mM Tris-HCl, pH 7.5, 250 mM NaCl, 1 mM DTT, 25 mM imidazole	Rap1b
E	25 mM Tris-HCl, pH 7.5, 250 mM NaCl, 1 mM DTT, 500 mM imidazole	Rap1b
F	25 mM Tris-HCl, pH 7.5, 250 mM NaCl, 1 mM DTT	Rap1b
G	25 mM Tris-HCl, pH 8.5, 500 mM NaCl, 20 mM imidazole	*Pa*IMPDH

2.3.2 Cell lysis and purification

The cell pellet was thawed using a 37 °C water bath and then resuspended with 7 mL of buffer (A, D or G; **Table 2.2**) supplemented with an EDTA-free protease inhibitor cocktail tablet (Roche) and 0.2 mg DNase I. The resuspended cells were then lysed using a One-Shot cell disruptor at a pressure of 30 kpsi before centrifugation at 37,500 rcf (Beckman Avanti J-25 with JA25.50 rotor) for 30 mins at 4 °C. After centrifugation, the supernatant containing the soluble fraction was harvested and the pellet discarded.

Affinity chromatography was used in the first step of purification. All over-expressed proteins studied in this thesis have a hexa-histidine N-terminal tag introduced from the pET15b-TEV vector (**Figure 2.1**). This His-tag has high affinity for divalent metals, thus the cell lysate was loaded onto a 5 mL HisTrap™ HP column (GE Healthcare) pre-charged with Ni^{2+} and equilibrated in buffer (A, D or G; **Table 2.2**). A linear concentration gradient of imidazole was applied to out-compete and thus elute the protein from the column using buffer (B or E; **Table 2.2**). Fractions containing protein were confirmed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and were pooled together before dialysis into buffer (C or F; **Table 2.2**) at room temperature.

Rap1b, *Bp*PBP3, *Pa*PBP3 and *Pa*IMPDH were purified with their His-tag still present. However *Sm*Lip was further purified by cleavage of the His-tag. This was carried out by incubating the protein with His-tagged TEV at 30 °C for 4 hrs. This protease recognises the amino acid sequence Glu-Asn-Leu-Tyr-Phe-Gln-Gly with the cleavage occurring between the Gln and Gly residues (**Figure 2.1**). The resulting mixture of protein/TEV was loaded on to a HisTrap™ column, which bound the cleaved His-tag, the His-tagged TEV and any uncleaved protein. The protein, from which the His-tag had been cleaved, was present in the flow-through peak, and fractions were confirmed by SDS-PAGE before pooling of the protein.

2.3.3 Size exclusion chromatography

Size exclusion chromatography was subsequently used to purify the protein further, separating proteins on the basis of their size; heavier proteins will pass through the column earlier than lighter proteins. Using a Superdex™ 200 26/60 column pre-equilibrated with buffer (C or F; **Table 2.2**) on a ÄKTA system (GE Healthcare), the protein was collected into fractions and confirmed pure by SDS-PAGE. The size exclusion column had previously been calibrated with molecular weight standards, blue dextran (> 2,000 kDa), thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29.5 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa); (GE Healthcare; data not shown).

Each fraction containing protein was then pooled into an Amicon 20 mL concentrator with an appropriate Mw cut-off and centrifuged at 3,500 rcf (Sigma 4K15 with a 373/E rotor).

2.4 Protein crystallography

2.4.1 Crystallisation of proteins

The crystallisation of proteins is based on the principles of thermodynamic equilibrium; the release of energy to achieve the minimal free energy state the system can reach (Rupp, 2010). This is the driving force of crystal growth. Protein molecules in a supersaturated solution, which is thermodynamically unstable, arrange themselves into an ordered lattice. This forms a crystalline aggregate and achieves a lower free energy state, which therefore brings about thermodynamic equilibrium in the system.

2.4.2 Crystallisation techniques

Two crystallisation methods were used in the studies presented in this thesis, the hanging drop method and the sitting drop method. Both methods use vapour diffusion

within a sealed environment to bring about crystallisation. This is achieved by placing a small mix of protein-precipitant alongside a reservoir of precipitant in a sealed environment. Due to the lower concentration of precipitant in the protein-precipitant mix, vapour diffusion occurs causing the net transfer of water to the reservoir until equilibrium is reached. The protein within the protein-precipitant mix becomes supersaturated and as a consequence, the protein is then forced out of solution and may form protein crystals.

The hanging drop method uses a small drop of protein, 0.5 μL - 2 μL, mixed with an equal volume of precipitant placed carefully on to a siliconised cover slip. The cover slip is then inverted and placed on top of a chamber containing 500 μL – 1 mL of precipitant. The edge of the well contains paraffin grease, which ensures the cover slip is secured and that the well is sealed from the external environment. This technique was used to optimise crystals.

The sitting drop method, as mentioned, uses vapour diffusion to achieve crystallisation of proteins. This technique uses a small drop of protein, 0.1 μL – 0.5 μL, mixed with an equal volume of precipitant. This drop is placed on a platform situated next to the reservoir containing 60 μL – 100 μL of precipitant. The well chamber is sealed from the external environment by an adhesive seal to allow vapour diffusion to take place. This technique was carried out in 96-well plates and is highly suited to screening a high number of crystallisation conditions using small volumes of protein. A Phoenix Liquid Handling System (Rigaku, Art Robins Instruments) was used to dispense the small volumes required for this technique.

2.4.3 Sparse matrix screening

To determine crystallisation conditions, a sparse matrix screen is performed which uses a large number of different conditions that have previously been used to crystallise other

proteins (Jancarik, 1991). The Classic, PEG, MPD and $AmSO_4$ suites (Qiagen) and the JCSG+ and PGA suites (Molecular Dimensions) were used, each containing 96 unique conditions. These conditions vary pH over a broad range (pH 4.0 – pH 10.5) and contain different salts and/or precipitants.

2.5 Crystal structure determination

2.5.1 The phase problem

Details on the theory and the phasing techniques outlined in this section can be obtained from Principles of X-ray Crystallography (Drenth and Mesters, 2006) and Biomolecular Crystallography (Rupp, 2010).

Briefly, in order to obtain the three-dimensional crystal structure of a protein, three parameters are required that define an electromagnetic wave; the wavelength (λ), the amplitude (A) and the phase (ϕ) (Taylor, 2010). The first two parameters are obtained from the diffraction experiment, however in doing such an experiment, all information about the phase is lost. Phase information of scattered X-rays cannot be directly measured. There are three principle methods of determining the phase for a new protein structure; multiple isomorphous replacement (MIR), multi/single-wavelength anomalous dispersion (M/SAD) and molecular replacement. During this study, SAD and molecular replacement phasing methods were used, and so the method of MIR will not be discussed further.

2.5.2 Patterson maps

A Patterson map can be constructed without any prior knowledge of phases and can be computed from any raw crystallographic data. This type of map contains interatomic distance information and is based on the Patterson function described as

$$P(u,v,w) = \frac{1}{V}\sum_{hkl} |F_{hkl}|^2 \, e^{-2\pi i(hu+kv+lw)}$$

The coordinates *u, v* and *w* locate a point in a Patterson map, as opposed to *x, y* and *z* that define a point within the unit cell. A contour map is produced that displays peaks for all interatomic distances, with the height of each peak a product of the corresponding electron density peak height. Patterson maps therefore can be used to solve the phase problem outlined in section 2.5.1 by locating heavy atoms in the unit cell of the protein, or by determining the orientation and translation of the search model during molecular replacement.

2.5.3 Single-wavelength anomalous dispersion (SAD)

Single-wavelength anomalous dispersion (SAD) is a method of obtaining phases based on the ability of elements to absorb X-rays and re-emit them with altered phases in a wavelength dependent manner (Taylor, 2010). This is called *anomalous scattering* or *anomalous dispersion*. The absorption of X-rays by heavy atoms invalidates Friedels law, meaning the reflections *hkl* are no longer equal in intensity to *-h-k-l*. This difference in symmetry-related reflections happens at all wavelengths, however it is more pronounced at the *absorption edge* of the heavy atom (Taylor, 2010).

This experimental phasing method was used to solve the structures of *Sm*Lip and Rap1b. Phases for *Sm*Lip were obtained using the iodides present in the crystallisation condition as a source of anomalous scattering. A quick halide soak (Dauter *et al*., 2000) with a high concentration of aqueous sodium iodide was used to obtained phases for Rap1b. Furthermore, several well-ordered sulfur atoms also provided a source of anomalous scattering. Anomalous scattering positions were identified using *PHENIX* (Adams *et al*., 2010) by Patterson methods as described in section 2.5.2 and phases were calculated using *PHASER* (McCoy *et al*., 2007). Model building was carried out using *RESOLVE* (Terwilliger, 2003).

2.5.4 Molecular replacement

Molecular replacement is a method of determining the crystal structure of a protein utilising another protein structure that has previously been solved (Rossmann and Blow, 1962). A protein with homologous amino acid sequence (and thus with similar fold), or containing functionally similar domains, can be used as a starting point to obtain new phases for a new structure. This typically requires more than approximately 25 % homology to succeed. The atomic structure factor phases depend on the location of the atoms in the unit cell, therefore the structure of the search model must be superimposed on to the target model before initial phases can be calculated. This process is separated into first finding the correct orientation and then the correct position for the search model.

The search for the correct orientation (rotation search) is achieved by calculating Patterson functions for both the search and target models. The Patterson function P_{calc} for the rotated search model is compared to the Patterson function P_{obs} of the target model through all possible rotations. The orientation is determined by the match of intramolecular Patterson vectors. If both models are in the same orientation, overlap of the Patterson maps will occur and summation of peaks will be produced. The search for the correct position (translation search) of the search model can be determined by matching intermolecular Patterson vectors, similar to the rotation search. Alternatively, comparing the expected structure factor amplitudes of the target model with the actual amplitudes from the diffraction data collected, can determine the correct position of the target. This is expressed as the R-factor and is carried out at all trial locations. The solution with the lowest R-factor is selected and initial electron density maps calculated. The resulting solution and maps are then inspected. This method of phasing was used for *Pa*IMPDH using the program *PHASER* (McCoy *et al*., 2007).

2.6 General software

PyMOL (www.pymol.org; Schrodinger, 2010) was used to create molecular structure figures in all instances. Sequence alignments were created using *T-coffee* (Di Tommaso *et al*., 2011) in the M-coffee mode (www.tcoffee.crg.cat) and were annotated using *ALINE* (Bond and Schüttelkopf, 2009). The software package Adobe Creative Suite 5 was used to edit, annotate and arrange figures accordingly.

2.7 Fragment screening

2.7.1 Reagents

Biotin-ligase (BirA) was made in-house by Martin Zoltner (University of Dundee) and provided ready to use at a concentration of 5 mg mL^{-1}. The following were purchased from the respective companies; biocytin (Sigma-Aldrich), 96-well and 384-well black plates (Matrix), "super" streptavidin biosensors (Fortebio), 96-well PCR plates (Thermo Scientific), adhesive PCR plate seals (4titude Ltd) and SYPRO Orange dye (Invitrogen).

The Maybridge Ro3 fragment library (Maybridge) was dissolved and stored in 100 % DMSO at a stock concentration of 40 mM. In addition, the Drug Discovery Unit (DDU, University of Dundee) fragment library, comprising 652 commercially available fragments, was dissolved and stored in 100 % DMSO at a stock concentration of 100 mM.

2.7.2 Biotinylation

*Pa*IMPDH and *Bp*PBP3 were cloned into a pET15b-TEV vector (see section 2.2.1) modified to include an N-terminal biotin acceptor peptide (BAP). Both proteins were expressed and purified as described in sections 2.2.2 and 2.3 respectively. Subsequently, 20 μM protein underwent biotinylation (10 mM Tris-HCl pH7.5, 200 mM KCl, 5 mM $MgCl_2$, 100 μM d-biotin, 500 μM ATP and 1 μM BirA) for 4 hrs at 30 °C.

The sample was then loaded onto a 5 mL HisTrap™ HP column (GE Healthcare) to remove His-tagged BirA. Free biotin was removed by size exclusion chromatography (see section 2.3.3) and the incorporation of biotin onto the protein was monitored by MALDI-TOF analysis (Proteomics Facility, University of Dundee). The protein was subsequently flash frozen in liquid nitrogen with the addition of 1 % glycerol and stored at -20 °C until required.

2.7.3 Differential scanning fluorimetry (DSF)

The stability of most proteins decreases with temperature; as the temperature increases, the protein begins to unfold and denature. The temperature at which the concentration of folded/unfolded protein is equal to one another is considered the melting temperature (T_m) of the protein. Differential scanning fluorimetry (DSF), or thermal-shift assay, utilises the fluorescent dye SYPRO orange in order to determine the T_m of a protein. This dye has affinity for hydrophobic residues, which become exposed upon protein denaturing and results in a change in fluorescence (**Figure 2.2**; Neisen *et al*., 2007; Pantoliano *et al*., 2001).

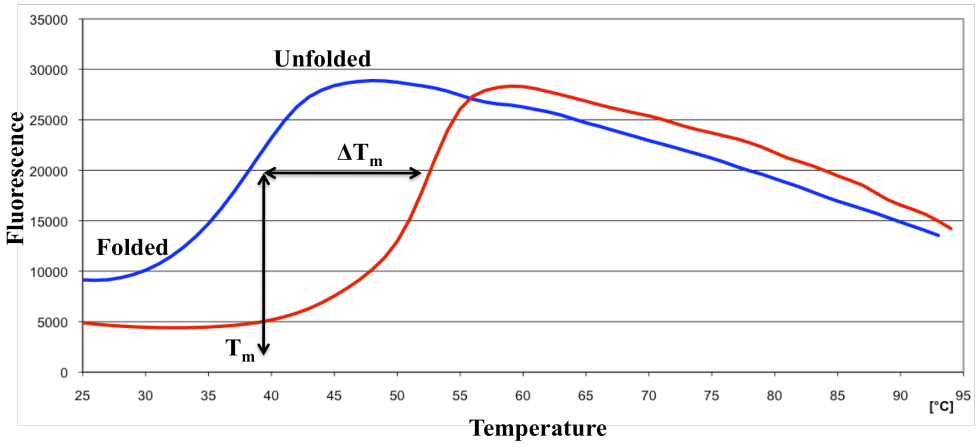


Figure 2.2 Melting curves of *Pa*PBP3.
The mid-point of the melting curve is considered the T_m value. The blue curve is *Pa*PBP3 and the red curve is *Pa*PBP3 in complex with cerbenicillin (Sainsbury *et al*., 2011).

This method can be used as a convenient way of determining whether a protein is folded or not. In addition, stabilising conditions can be determined in order to improve purification and crystallisation of a protein. This is achieved by comparing the T_m values for the protein in the absence and presence of the stabilising condition, with a positive difference (ΔT_m) an indication that the protein is more stable. Furthermore, potential cofactors, compounds or fragments can be screened for a protein using the same principle, with a positive ΔT_m suggesting potential binding with the protein. A negative ΔT_m may also indicate potential binding, however this is likely to be unspecific.

Fragment screening with the Maybridge Ro3 library was carried out with *Pa*IMPDH, *Bp*PBP3 and *Pa*PBP3. Solutions of protein at 10 μM with 1000x SYPRO orange were made and 39 μL dispensed into a 96-well thin-wall PCR plate prior to adding 1 μL of fragment to each well. The plates were then sealed with adhesive PCR seal and heated in an Mx3005p qPCR machine (Stratagene) from 25 °C to 95 °C at a rate of 1 °C min^{-1}. Fluorescence changes were monitored with excitation and emission wavelengths of 492 nm and 610 nm respectively. The change in fluorescence was plotted against temperature, producing a melting curve for the protein (Niesen *et al*., 2007). Reference wells (i.e. wells consisting only of protein and dye) were used to compare T_m values. In order to rule out any effects caused by the presence of the buffer the protein is stored in, wells with buffer and dye were used as controls. Experiments were carried out in triplicate and nonlinear regression analysis was determined using the curve-fitting program Prism (GraphPad Software). A Microsoft Excel script was used for convenient observations and comparison of melting curves and thermal shifts (ftp://ftp.sgc.ox.ac.uk/pub/biophysics). The fragments were screened at a final concentration of 1 mM.

2.7.4 Surface plasmon resonance (SPR)

Surface plasmon resonance (SPR) is an optical analytical technique used to detect and characterise biomolecular interactions (Hoa *et al*., 2007). When monochromatic polarised light is applied to the gold surface of a sensor chip, the intensity of the reflected light is reduced at a specific angle called the SPR angle. This angle depends on the refractive index at the surface of the sensor chip (**Figure 2.3A**). Immobilised protein (the ligand) on the surface of the sensor chip interacts with binding partners (the analyte) and an increase in the refractive index occurs, resulting in an altered SPR angle (**Figure 2.3B**). The change in the SPR angle is monitored by the reduced intensity of the reflected light and is measured in real time. The result is plotted as resonance units (RU) against time. The sensitivity of this technique is highly suited for studying the interaction between proteins and weak binding fragments (Hoa *et al*., 2007).

Biotinylated *Bp*PBP3 was immobilised onto a streptavidin sensor (GE Healthcare) and screened against the DDU fragment library by Dr. Leah Torrie (University of Dundee). The final concentration of fragments used during the screening was 150 μM.

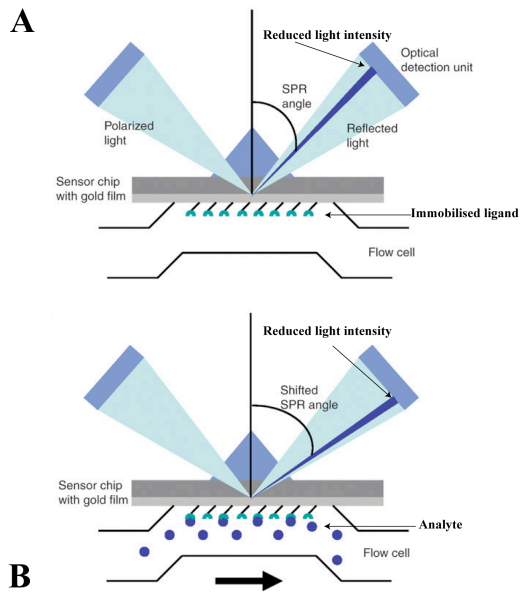


Figure 2.3 Principles of SPR.
Polarised light is reflected from the surface of the sensor chip. The intensity of the reflected light is reduced at the SPR angle (**A**). The interaction between ligand and analyte results in a change in the refractive index causing a change in the SPR angle (**B**). This change is monitored in real time. Figure edited from (Madeira *et al*., 2009).

2.7.5 Bio-layer interferometry (BLI)

Bio-layer interferometry (BLI) is a label free optical analytical technique used to analyse biomolecular interactions. This technique monitors in real time the interference pattern of white light reflected from two surfaces within a biosensor (Wartchow *et al*., 2011). The first surface is an internal reference layer, while the second surface is a layer of immobilised protein (**Figure 2.4A**). As the immobilised protein on the biosensor tip binds to the analyte in solution (**Figure 2.4B**), an increase in optical density occurs resulting in a shift in wavelengths (**Figure 2.4C**).

BLI was used to screen biotinylated *Pa*IMPDH and *Bp*PBP3 against the DDU fragment library, using streptavidin biosensors to immobilise protein at 25 μg mL^{-1} for 10 mins. Biocytin at 10 μg mL^{-1} was then used to block any remaining free streptavidin binding sites for 60 sec prior to washing and equilibrating the biosensors in buffer for 5 mins. The fragments were screened at a final concentration of 200 μM in 384-well format using an Octet Red system (Fortebio) in 16-channel mode. The temperature of the chamber was set to 25 °C.

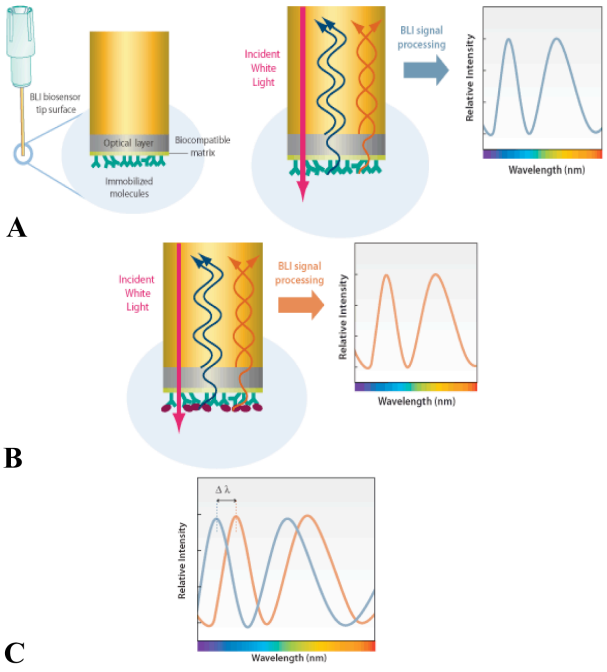


Figure 2.4 Bio-layer interferometry (BLI).
Light is applied down a biosensor and an internal reference (**A**) is used to compare the wavelength of the light generated upon the binding of the analyte (**B**). This binding results in a shift in wavelengths (**C**) from the change in optical density at the biosensor surface. Figures obtained from www.fortebio.com.

2.8 Enzyme kinetics

The enzymatic characterisation of *Pa*IMPDH was determined spectrophotometrically using a Shimadzu UV-2450 to monitor the formation of NADH at 340 nm. This is based on the assay described previously for characterising human IMPDH (Carr *et al*., 1993; Miller and Adamczyk, 1976). Assays for determining the K_m and V_{max} values of the substrate, inosine 5'-monophosphate (IMP), were performed at 30 °C in 100 mM Tris-HCl pH 8.0, 100 mM KCl, 3 mM EDTA, 1 μM *Pa*IMPDH and between 500 μM and 25 mM IMP, with the concentration of NAD^+ fixed at 5 mM. Conversely, assays for determining the K_m and V_{max} values of the cofactor NAD^+ were performed between 10 μM and 5 mM, with the IMP concentration fixed at 25 mM. Assays were carried out in triplicate in 1 ml volumes in 1 cm path-length acrylic cuvettes (Sarstedt). Reactions were initiated with the addition of IMP and the increase in absorbance of NADH monitored continuously for 5 mins.

The uncompetitive inhibitor mycophenolic acid (MPA; Allison and Eugui, 2000) was dissolved in 100 % DMSO and the K_i value was calculated using 1 μM and 10 μM MPA, with the concentrations of IMP being the same as was used to calculate the substrate K_m. The concentration of NAD^+ was fixed at 5 mM. Single point percentage inhibition was measured with the identified fragments. As a control, the additions of 0.1 %, 1 % and 10 % DMSO to enzyme reactions were compared to reactions in the absence of DMSO. The activity of the enzyme was comparable; therefore the final concentration of DMSO in experimental reactions were set to 1 % or less. Analyses of the kinetic data were carried out using the software Prism (GraphPad Software).

CHAPTER 3

RESULTS AND DISCUSSION: *Sm*Lip

3.1 Aims

The aim was to structurally characterise recombinant *S. marcescens* Lip (*Sm*Lip). A conserved component of the type VI secretion system (T6SS; Boyer e*t al*., 2009), *Sm*Lip is an essential lipoprotein (Aschtgen *et al*., 2008; Murdoch *et al*., 2011), although its function within the T6SS is not fully understood. The crystal structure would allow us to compare with protein structures deposited in the PDB (Berman *et al*., 2002) in order to infer its possible function within the T6SS.

3.2 Construction of the pET15b-TEV-*Sm*Lip vector

A truncated form of the *S. marcescens lip* gene encoding the amino acid residues 30 – 176 was PCR amplified from genomic DNA (see section 2.2.1). Secondary structure prediction (*XtalPred*; Slabinski *et al*., 2007) and signal peptide prediction (*SignalP*; Petersen *et al*., 2011) were used as a guide for primer design and the following primers were used; forward primer 5' – catatgGCCAAAAGCGTGCCGTCGCGTTACAG and reverse primer 5' – ggatccTCAGTCGACCTTTTTTACGGGGCGCAGGC (lower case sequences correspond to the *Nde*I/*Bam*HI restriction sites used for cloning). The PCR parameters are described in section 2.2.1.

A single DNA band, corresponding to around 400 bp, was observed on a 1 % agarose gel (**Figure 3.1A**). The band was excised from the gel and DNA was extracted using a Qiagen Gel Extraction kit. The PCR product was ligated into the pCR-Blunt II TOPO® vector using the Zero Blunt® TOPO® Cloning Kit (Invitrogen). The *Sm*Lip gene was subsequently cut out of the pCR-Blunt II TOPO® vector using the *Nde*I/*BamH*I restriction enzymes (**Figure 3.1B**) at 37 °C for 1 hr. The *Sm*Lip insert was then ligated into the pET15b-TEV cloning vector using T4 DNA ligase (NEB).

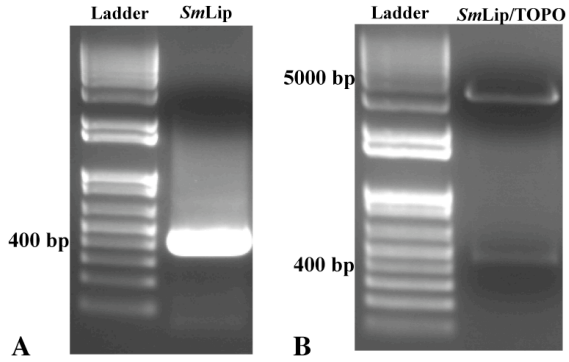


Figure 3.1 Cloning of *Sm*Lip.
A) The amplified PCR product is shown corresponding to approximately the correct size of the gene. **B**) *Sm*Lip (lower band) is cut out of the PCR-BluntII-TOPO® vector (top band).

3.3 Recombinant *Sm*Lip expression and purification

The *Sm*Lip pET15b-TEV plasmid was transformed into *E. coli* Bl21 (DE3) pLysS cells (Stratagene). Cultures were grown in AI media (Studier, 2005) supplemented with 50 μg L^{-1} carbenicillin as described in section 2.2.2.

Cell lysis was carried out as described in section 2.3.2. The supernantant containing soluble protein was harvested and loaded on to a 5 mL HisTrap™ HP column (GE Healthcare) pre-charged with Ni^{2+}. A large peak at around 300 mM imidazole was obtained and appropriate fractions were pooled and incubated with His-tagged TEV protease to remove the His-tag on *Sm*Lip. The resulting mixture was passed down a second 5 mL HisTrap™ HP column and the flow through fractions containing the cleaved protein were pooled (see section 2.3.2). The protein was loaded onto a calibrated size exclusion column (see section 2.3.3) where a single peak corresponding to a monomer was observed. Samples were taken for SDS-PAGE analysis (**Figure 3.2**) and appropriate fractions were pooled and concentrated to 10 mg mL^{-1}, with a yield of 2.5 mg L^{-1} of bacterial culture obtained. A theoretical extinction coefficient of 16960 M^{-1} cm^{-1} at 280 nm, was used to estimate protein concentration (*PROTPARAM*; Gasteiger *et al*., 2005); the theoretical mass of one subunit is estimated as 16.1 kDa with a calculated isoelectric point (pI) of 5.4. The purified protein sample was stored at 4 °C until further use.

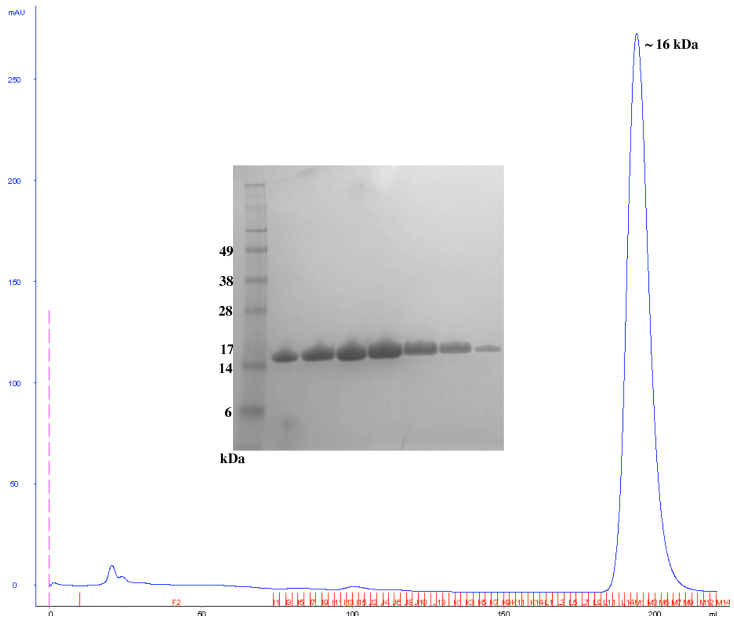


Figure 3.2 Size exclusion chromatography.
The absorbance (y-axis) at 280 nm (mAU) is plotted against eluted volume (ml) on the x-axis. A single absorbance peak after approximately 190 ml eluted volume was observed, consistent with a retention volume for a protein of around 16 kDa. 4 ml fractions were collected.

3.4 Crystallisation and data processing

Crystallisation screening with the Classic, PEG, MPD and $AmSO_4$ suites (Qiagen) and the JCSG+, PGA (Molecular Dimensions) suites were set up initially with 10 mg mL^{-1} of protein in 25 mM Tris-HCl pH 7.5 250 mM NaCl at 20 °C (see section 2.4.3). This was carried out using the sitting drop vapour diffusion method in 96-well plates using the Phoenix Liquid Handling System (Rigaku, Art Robins Instruments) (see section 2.4.2). Multiple crystals grew in several crystallisation conditions after approximately two days (**Figure 3.3A**), however only two conditions produced diffraction quality crystals. The conditions were; 20 % polyethylene glycol (PEG) 3350, 200 mM KI and 15 % PEG 3350, 200 mM NaCl (**Figure 3.3B**).

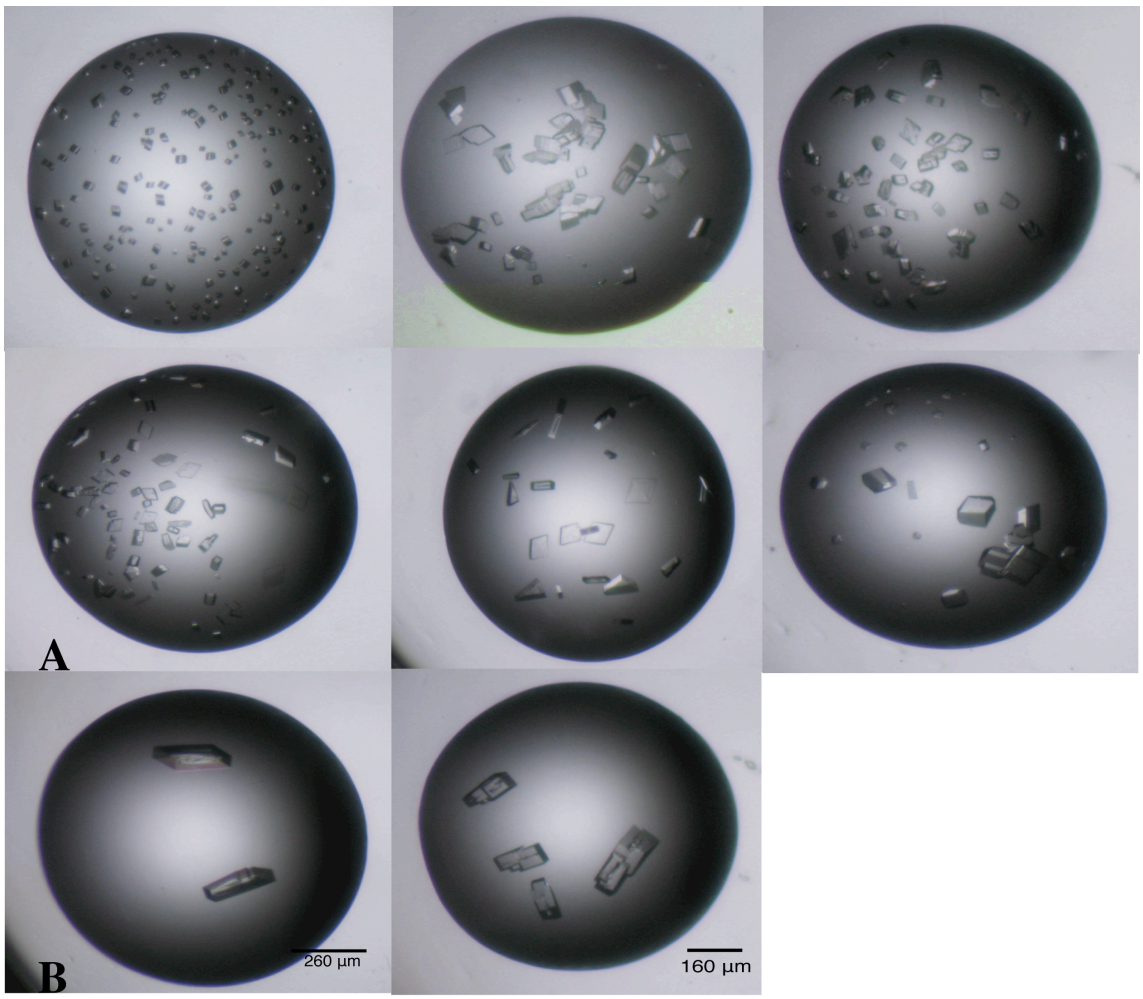


Figure 3.3 Crystals of *Sm*Lip.
A) Mulitple crystals of *Sm*Lip grew after approximately two days in several conditions **B**) Diffracting crystals in; 20 % polyethylene glycol (PEG) 3350 200 mM KI (left) and 15 % PEG 3350 200 mM NaCl (right).

Crystals from the iodide containing condition were optimised (**Figure 3.4A**) and then characterised. Dataset I was measured in-house using a Rigaku MicroMax-007 rotating-anode X-ray generator (copper K_α, λ=1.5418 Å) coupled to an R-AXIS IV^{++} image plate detector (**Figure 3.4B**). A single crystal from the chloride condition was stored in liquid N_2 then subsequently used to measure a 1.92 Å high-resolution dataset (dataset II) at the European Synchrotron Radiation Facility (ESRF; Grenoble, France) on beam line ID29 using an ADSC Q315R detector. The addition of 10 % glycerol proved to be a suitable cryoprotectant for both conditions. Data were indexed and integrated using *XDS* (Kabsch, 2010), with space group *C*2 and unit cell lengths *a*= 139.7 *b*= 77.8 Å and *c*= 54.5 Å (**Table 3.1**). Data were scaled and the space group

suggested using *SCALA* and *POINTLESS* (Evans, 2006) from the *CCP4* program suite (Winn *et al*., 2011).

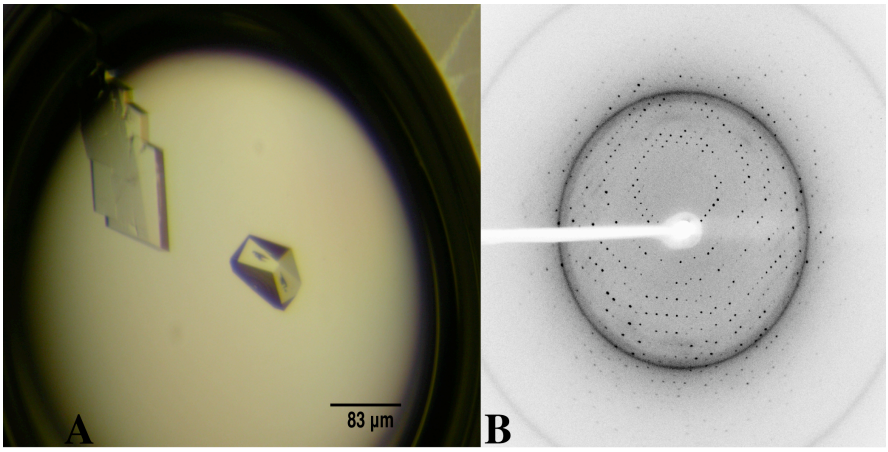


Figure 3.4 Crystal optimisation.
A) Iodide containing condition optimised to 25 % PEG 3350 300 mM KI. **B**) In-house diffraction to 2.3 Å produced from optimised crystals.

3.5 Structure determination of *Sm*Lip

3.5.1 Phasing

A search of the PDB using the amino acid sequence of *Sm*Lip indicated no similarity to any structures deposited. Dataset I was used to solve the structure of *Sm*Lip by SAD methods targeting the iodides present in the crystallisation condition (Dauter *et al*., 2000). SAD is a method of obtaining phases based on the ability for elements, in this case iodide, to absorb X-rays (as well as emit them) in a wavelength dependent manner (Taylor, 2010; see section 2.5.3). Thirteen potential iodide positions were identified and phases were calculated, giving a figure-of-merit of 0.43 to 2.35 Å. Subsequently twelve of these positions were confirmed by refinement with this dataset (**Figure 3.5**). Density modification was carried out using solvent flattening, histogram matching and averaging on the basis of non-crystallographic symmetry (NCS). The resulting electron density clearly revealed the molecular structure (**Figure 3.5**).

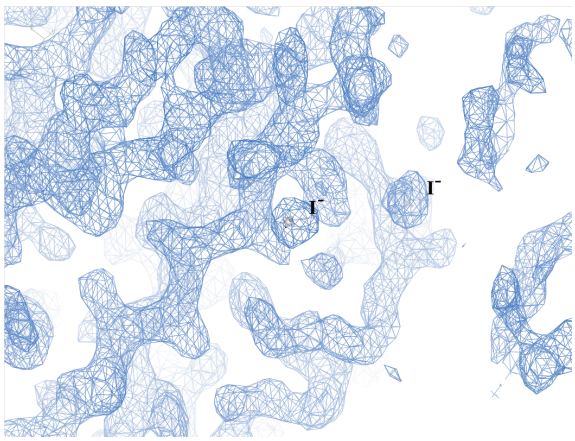


Figure 3.5 Density modified map. Several iodides from the crystallisation condition were identified and used for SAD phasing. Map contoured at 1.5 σ.

3.5.2 Model building and refinement

The initial model consisted of 293 residues, with a correlation coefficient of 0.55 and R_{work} and R_{free} values of 46 % and 49 % respectively. The first round of model building, carried out manually in *COOT* (Emsley *et al*., 2010), extended this to 467 residues, with a correlation coefficient of 0.72 and R_{work} and R_{free} values of 33 % and 37 % respectively. NCS restraints were employed in the initial refinement calculations, which were performed using *REFMAC5* (Murshudov *et al*., 2011). At this point the high-resolution synchrotron data (dataset II, 1.92 Å resolution) was used to continue the analysis. The refinement proceeded with the release of NCS restraints, the incorporation of water molecules, a Na^+ ion, ethylene glycol and a number of side chains with dual rotamer conformations. This dataset was derived from crystals grown in the presence of chloride instead of iodide. However, we did not assign any chloride ions noting that typical water molecules occupy the iodide binding sites previously identified. The refinement was terminated when there were no significant changes in R_{work} and R_{free} and inspection of the difference density map suggested that no further corrections or additions were justified. Crystallographic statistics are given in **Table 3.1**.

Table 3.1 Crystallographic statistics.

	Dataset I	**Dataset II**
Space group / wavelength (Å)	*C*2 / 1.5418	*C*2 / 1.007
a, b, c (Å), β (°)	139.7, 77.6, 54.3, 98.4	139.7, 77.8, 54.5, 98.3
Resolution[a] (Å)	19.7 – 2.35 (2.48 – 2.35)	39.8 – 1.92 (2.02 – 1.92)
No. reflections recorded	94650 (11192)	280802 (38597)
Unique reflections	23495 (3109)	43478 (6188)
Completeness (%)	98.3 (90.0)	98.2 (95.8)
Multiplicity / $\langle I/\sigma I \rangle$	4.0 (3.6) / 30.9 (6.6)	6.5 (6.2) / 20.6 (3.7)
Anomalous completeness (%)	95.7 (84.5)	-
Anomalous redundancy	2.0 (1.8)	-
Wilson B (Å^2)	47.7	32.7
Residues / waters	-	541 / 336
R_{merge}[b] (%)	2.6 (17.1)	4.9 (44.2)
R_{work}[c], R_{free}[d] (%)	-	22.0 / 29.2
Ave. *B*-factor (Å^2)		
Chain A, B, C, D	-	36.1, 41.6, 49.5, 59.8
Waters, Na^+, ethylene glycol		46.2, 37.8, 60.2
Cruickshank DPI[e] (Å)	-	0.2
Ramachandran plot		
Most favoured	-	516 residues
Additional allowed	-	21 residues
Outliers	-	Molecule D: Phe97, Asp129; molecule B: Pro142, Ser154
RMSD on ideal values[f]		
Bond lengths (Å) / angles (°)	-	0.01 / 1.42

a. Values in parentheses refer to the highest resolution shell. **b.** $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. **c.** $R_{work} = \sum_{hkl}||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. **d.** R_{free} is the same as R_{cryst} except calculated with a subset, 5 %, of data that are excluded from the refinement calculations. **e.** Diffraction Precision Index (Cruickshank, 1999). **f.** (Engh and Huber, 1991).

3.6 The *Sm*Lip structure

3.6.1 Overall structure

The asymmetric unit consists of four polypeptide chains, labeled A-D, with an estimated solvent content of 45 % and a V_m (Matthews, 1968) of 2.27 $Å^3$ Da^{-1}. Disorder was evident at several positions where it was not possible to interpret diffuse and weak electron density. Consequently several residues are absent from the model. Molecule A consists of residues 32 to 173, molecule B 33-142, 147-176; molecule C 34-50, 53-143 and 147-175; molecule D 33-50, 55-175. The geometry and quality of the model was validated using *MOLPROBITY* (Chen *et al*., 2010) and the subsequent Ramachandran plot reveals 99.3 % of residues were found in the allowed region. There were four residues in the disallowed region with weak electron density and exhibiting above average B-factors (**Table 3.1**). The four *Sm*Lip polypeptide chains are similar with the root-mean-square deviations (RMSD) between superimposed C_α atoms ranging from 1.3 Å (monomers A and B) to 0.8 Å (monomers A and D) with an average value of 0.95 Å.

The structure of *Sm*Lip is dominated by an eight-stranded β-sandwich with strands in the following order; 8-7-1-4 and 6-5-2-3 (**Figure 3.6A**). In addition, there are three short α-helical segments and three 3_{10}-helix turns, two of which are on a loop that extends away from the core of the protein (**Figure 3.6B**). This loop is located in close proximity to the N-terminus, which is anchored to the outer membrane via the lipidated Cys26 (Bos *et al*., 2007; Tokuda and Matsuyama, 2004). It is possible this loop region makes interactions with the outer membrane (**Figure 3.6B**). The two 3_{10}-helix turns could potentially distort to form an α-helix and embed into the membrane. This would increase stabilisation of the protein and perhaps ensure the protein is positioned correctly to form protein-protein interactions with other components within the T6SS. In addition, there are several other exposed loop regions that can be seen in the crystal

structure, with loop 1 protruding out from the core of the protein, potentially making interactions with other proteins in the periplasm.

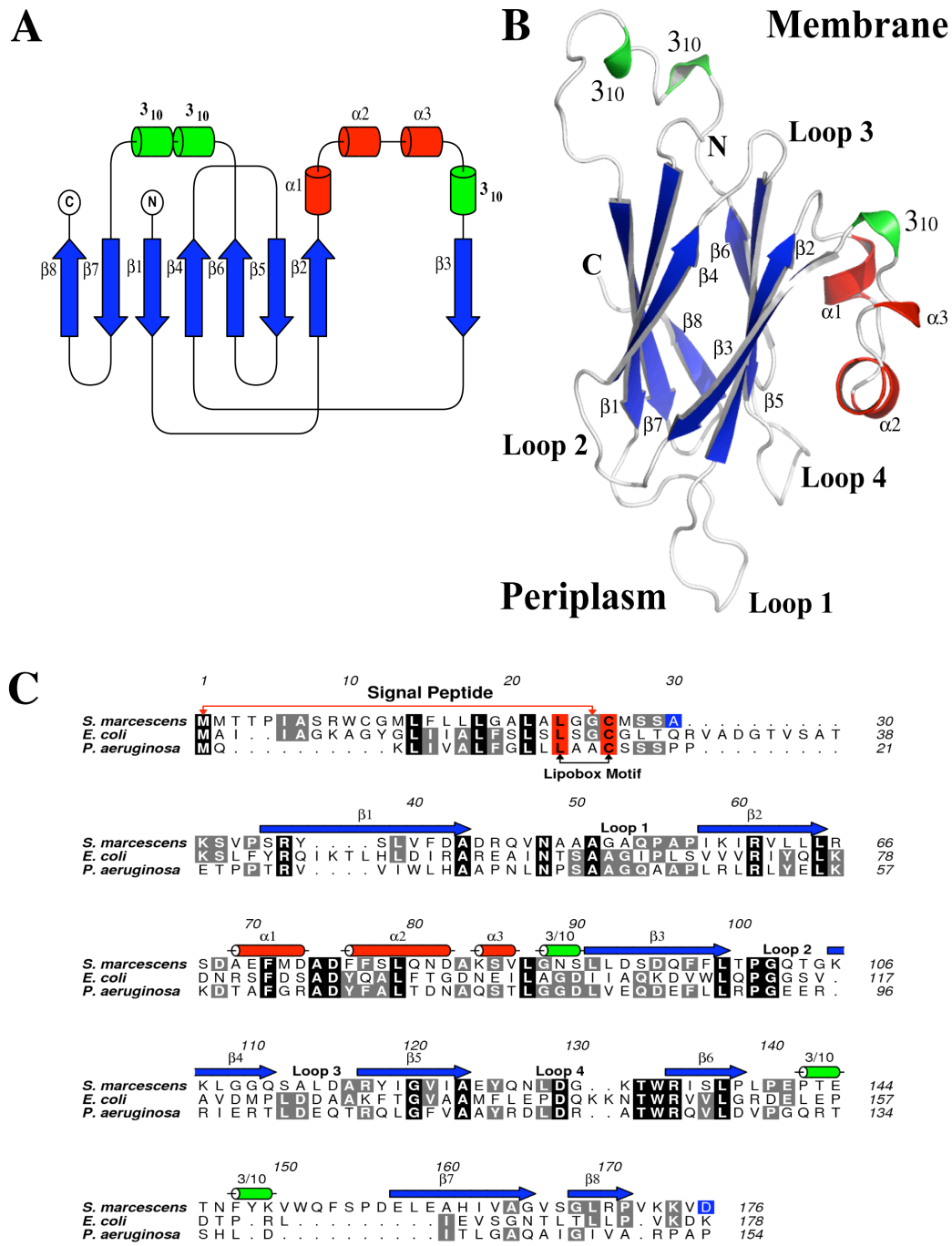


Figure 3.6 The structure of *Sm*Lip.
A) The topology of *Sm*Lip shows the arrangement of the secondary structure, with β sheets coloured blue, α-helices coloured red and 3_{10} helices coloured green. Image prepared using TopDraw (Bond, 2003). **B**) The structure of *Sm*Lip (PDB 4A1R) coloured using the same colour scheme as **A**. The structure is orientated with the N-terminus pointing towards the outer-membrane, where the excluded lipidated cysteine would anchor the protein into the membrane. **C**) Sequence alignment of *Sm*Lip with the homologous proteins from enteroaggregative *E. coli* (GenBank CBG37366.1) and *P. aeruginosa* (NCBI Reference Sequence NP_248770.1, PA0080). The secondary structure is depicted using the same colour scheme as **A**.

Despite *Sm*Lip being a conserved component of the T6SS (Boyer *et al*., 2009), there is a large degree of variation in the amino-acid sequences of homologous proteins. This is highlighted in the sequence alignment with homologues (**Figure 3.6C**), where *Sm*Lip shares approximately 20 % sequence identity with enteroaggregative *E. coli* (EAEC) Lip and nearly 40 % sequence identity with *P. aeruginosa* Lip. The conservation of a number of residues is noted in the region corresponding to loop 1 and loop 2, and at the secondary structure elements α1, α2 and β6 (**Figure 3.6C**). The side chain of Asn48 in loop 1 accepts a hydrogen bond from the main-chain amide of Gln126 in loop 4 and aids in defining the conformation of loop 4. The conserved region at loop 2 (Leu99-X-Pro101-Gly102) defines the structure of the turn after β3 and forms the loop (data not shown).

3.6.2 Analysis of the quaternary structure

Size exclusion chromatography indicates a monomer in solution for *Sm*Lip. However, four polypeptides are present in the asymmetric unit and their interactions were analysed using the *PISA* web service (Krissinel and Henrick, 2007). The accessible surface area (ASA) of the polypeptides range from 8,200 $Å^2$ for molecule D to 8,510 $Å^2$ for molecule A. Each molecule in the asymmetric unit interacts with two of the other three and two types of protein-protein interface are formed between molecules A-B, C-D (interface I) and A-C, B-D (interface II; **Figure 3.7A**). The largest interface is the type I interface and covers an area that is approximately 13 % of the ASA of the *Sm*Lip molecule. This coverage would indicate a potential for a biologically relevant dimer (Krissinel and Henrick, 2007). This interface is primarily formed by the anti-parallel alignment of two β7 strands (**Figure 3.7B**). Three aromatic residues, Phe147, Trp151 and Phe153 contribute van der Waals interactions to the association and by virtue of their relative bulk also to the ASA. The type II interface covers approximately 6.5 % of

the ASA of a molecule, a level typical of values observed simply due to molecular packing in a crystal lattice. This interface is formed by the anti-parallel alignment of two β4 strands (**Figure 3.7B**).

Evidence against *Sm*Lip forming a tetramer in the cell can be found upon closer inspection of the amino acid sequence and the crystal structure. Analysis of the sequence of *Sm*Lip indicated the presence of a signal peptide (Petersen *et al*., 2011), and the lipidation of Cys26 at the N-terminus. This linkage anchors the protein to the membrane (Bos *et al*., 2007; Tokuda and Matsuyama, 2004). The spatial placement of the N-terminal residue of each polypeptide chain in the asymmetric unit suggests it is unlikely that an oligomeric assembly could form when the protein is anchored to the membrane. The N-termini of molecules A and D are on the same side of the tetrameric assembly, but are opposite to those of molecules B and C. As explained, there are only two interface types observed and there are no direct interactions formed between molecules A and D or molecules B and C, making the structure incompatible with an oligomeric assembly when located at the membrane. Bacterial two-hybrid studies indicate *Sm*Lip undergoes self-interaction, with this system unable to distinguish between dimerisation or higher order oligomerisation (Rao *et al*., 2011). This is consistent with the crystal structure showing a tetrameric assembly with a possible dimer. However, as discussed, this arrangement is incompatible with a membrane-bound protein. This may be due to different concentrations and experimental conditions used. It is therefore more likely that *Sm*Lip is a membrane-bound monomer that functions by protein-protein interactions. *Sm*Lip is therefore able to undergo self interaction when unbound to the membrane, as observed from bacterial two-hybrid results (Rao *et al*., 2011).

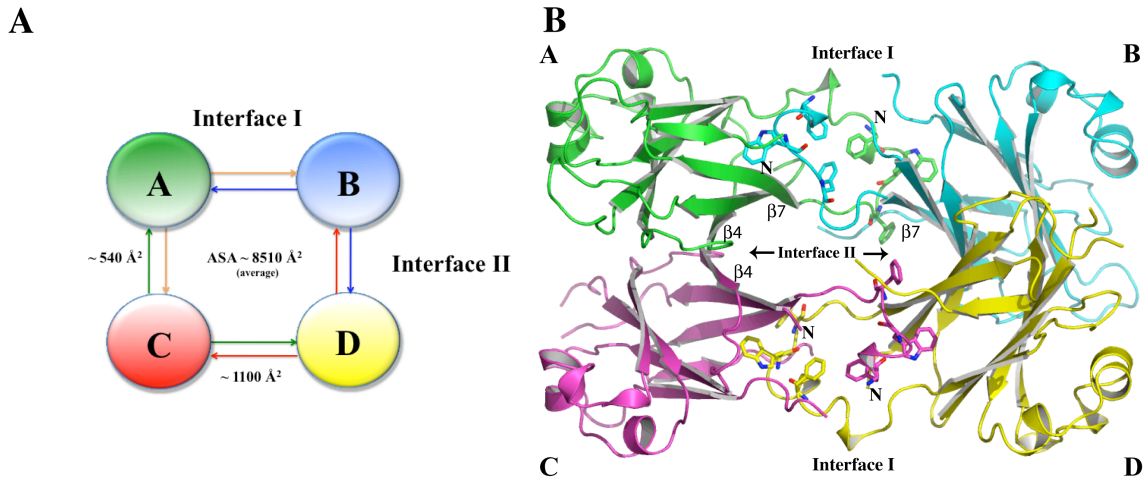


Figure 3.7 Crystallographic interfaces.
A) Schematic diagram of the two different interfaces observed in the crystal structure of *Sm*Lip and the ASA of each interface. **B**) The asymmetric unit, with the two interfaces labelled. Residues that contribute significantly to the type I interface (Phe147, Trp151 and Phe153) are depicted as sticks.

3.6.3 Comparisons with structural homologues

To infer the possible function of *Sm*Lip, a search for structural homologues using *PDBeFold* (http://pdbe.org/fold) was carried out. The closest match gave a Z-score of 6.1 for sea bream transthyretin (PDB 1SN0; Eneqvist *et al*., 2004). Superimposing the two structures matches 84 residues with an RMSD of 2.7 Å (**Figure 3.8A**). The β-strands align well, however transthyretin lacks the small α-helical domain present in *Sm*Lip and the loop regions fail to align (**Figure 3.8A**). The sequence identity between the two proteins is approximately 7 %. A clear structural relationship can be discerned, establishing *Sm*Lip as a new member of the transthyretin-like protein family. The only other member of this protein family is 5-hydroxyisourate hydrolase, an enzyme found in prokaryotes (EC 3.5.2.17; Hennebry *et al*., 2006), leading to the conclusion that this family represents an example of divergent evolution (Hennebry, 2009). The sequence identity shared with *Sm*Lip is 6 %, however a similarity in fold is still evident (**Figure 3.8B**). It is unlikely *Sm*Lip has any hydrolase activity as the residues important for the

catalytic function of 5-hydroxyisourate hydrolase (His7, Arg41, His92 and Ser108) are absent.

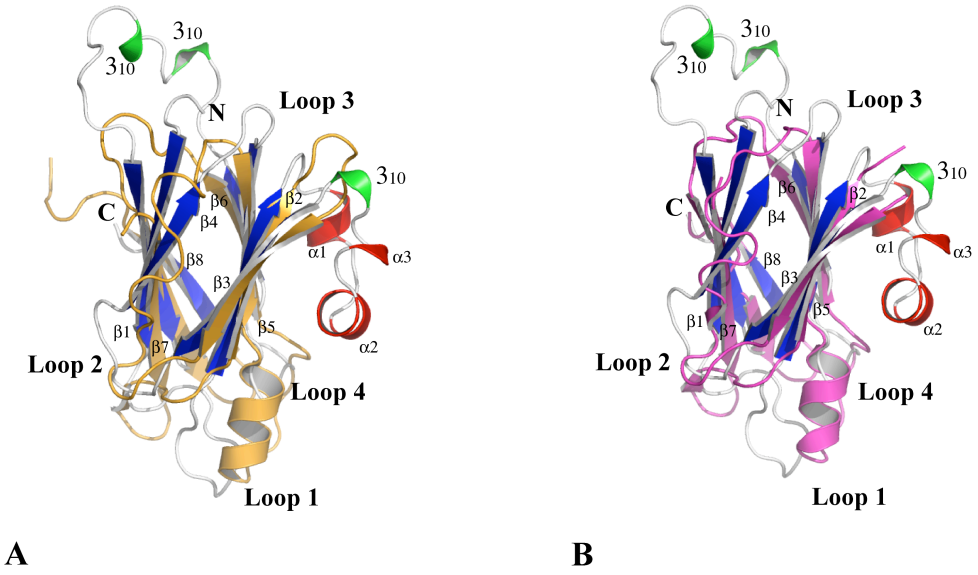


Figure 3.8 Comparison with structural homologues.
A) A clear structural relationship can be seen when *Sm*Lip is superimposed onto sea bream transthyretin (PDB 1SN0; orange). **B**) *Sm*Lip is also structurally similar to 5-hydroxyisourate hydrolase (PDB 2GPZ; magenta), another member of this protein family. *Sm*Lip is coloured using the same colour scheme as in **Figure 3.6A**.

Transthyretin is a tetrameric protein involved in binding the hormone thyroxine and forming a complex with retinol-binding protein (Prapunpoj and Leelawatwattana, 2009). The interface of the tetramer is formed by two of the eight β-sheets present (PDB 2PAB; Blake *et al.*, 1978). When superimposing the structure of *Sm*Lip onto transthyretin, one polypeptide overlays onto one transthyretin polypeptide. Furthermore, an overlay of *Sm*Lip with transthyretin places the thyroxine-binding site on the surface of *Sm*Lip (**Figure 3.9**). This is in contrast to the thyroxine-binding site in transthyretin, where binding occurs at the tetramer interface. This would suggest that it is unlikely that *Sm*Lip acts to bind hydrophobic ligands of the type that transthyretin can bind.

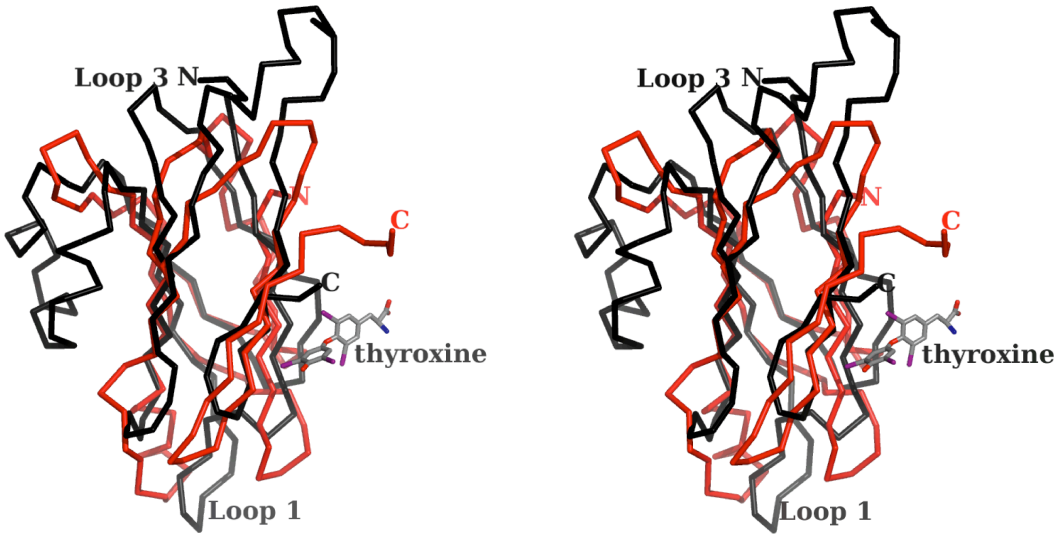


Figure 3.9 Stereo image of an overlay of *Sm*Lip and transthyretin in complex with thyroxine. The polypeptide trace for *Sm*Lip is black, transthyretin red. Thyroxine is shown as sticks coloured purple for I, grey C, red O and blue N.

3.7 Discussion

The T6SS is implicated in a number of different phenotypes including the virulence of *B. pseudomallei* towards mice (Chen *et al*., 2011) and the antibacterial activity of *S. marcescens* (Murdoch *et al*., 2011; see section 1.2). Recently, a number of crystal structures of components of the T6SS have improved our understanding of the function and assembly of these components within the secretion apparatus (**Figure 3.10**). This has subsequently led to the proposal that the T6SS may function in a similar manner to that of an inverted bacteriophage (Bönemann *et al*., 2010; Cascales and Cambillau, 2012; Records, 2011; see section 1.2).

The lipoprotein *Sm*Lip is an essential component of the T6SS, with deletion of this gene resulting in a non-functional T6SS (Aschtgen *et al*., 2008; Murdoch *et al*., 2011). The amino acid sequence of *Sm*Lip shows the presence of a signal peptide, with cleavage by signal peptidase II occurring between Gly25 and Cys26 (**Figure 3.6C**; Petersen *et al*., 2011). The residue after the lipidated Cys26, Met27 (+2 position), suggests that this protein is exported to the outer membrane via the Lol system (Bos *et al*., 2007; Tokuda and Matsuyama, 2004). The localisation of the EAEC Lip homologue

(*Ec*TssJ) has previously been confirmed (Aschtgen *et al*., 2008), while the localisation of *Sm*Lip was confirmed using anti-Lip polyclonal antibodies to probe the major cell fractions (Rao *et al*., 2011).

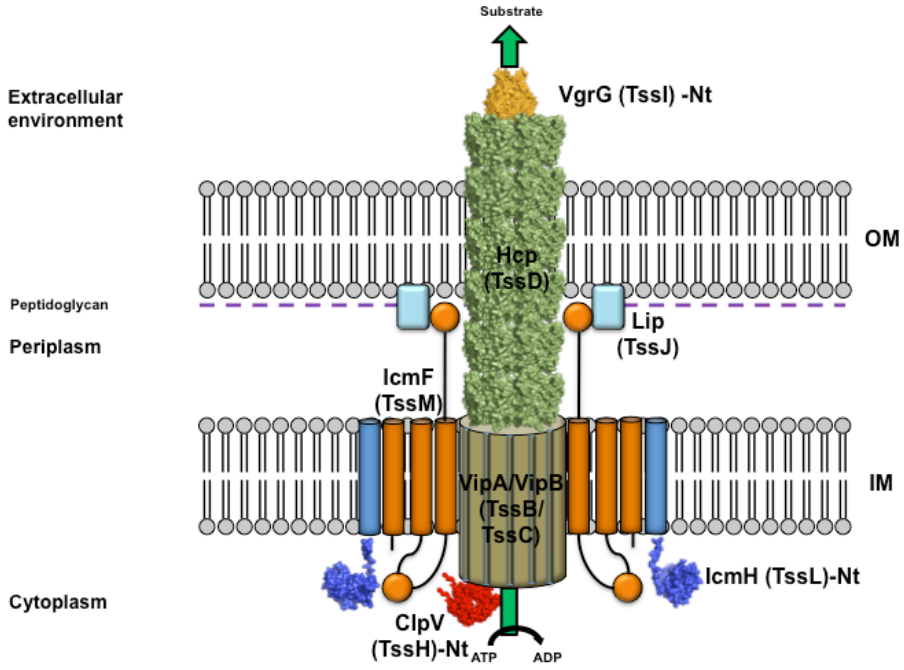


Figure 3.10 The T6SS.
Two distinct sub-assemblies have now been identified; one resembling contractile bacteriophage tail and the other a membrane spanning complex that anchors the first sub-assembly to the cell envelope. PDB; 1Y12, 2P5Z, 3U66 and 3ZRI. Figure adapted from (Bönemann *et al*., 2010; Cascales and Cambillau, 2012).

During the course of these studies, the *Ec*TssJ crystal structure became available (PDB 3RX9; Felisberto-Rodrigues *et al*., 2011). The authors note the same structural relationship with transthyretin that was observed with the *Sm*Lip structure (Rao *et al*., 2011). Superimposing the *Ec*TssJ structure with *Sm*Lip matches 120 residues with an RMSD of 1.5 Å (**Figure 3.11A**) and sequence identity of approximately 20 % (**Figure 3.11B**). A noticeable difference between the two structures is the region comprising the two 3_{10}-helices between β6 and β7 in *Sm*Lip. While this region corresponds to a large loop of 18 residues that extends away from the core of *Sm*Lip, the *Ec*TssJ structure comprises just a single 3_{10}-helix corresponding to only 9 residues. This difference

results in the *Ec*TssJ structure being more compact and globular than that of *Sm*Lip. The significance of this, if any, has yet to be determined. However it is worth noting that this domain is involved in the type I interface between molecules A-B and C-D in the *Sm*Lip asymmetric unit (**Figure 3.7B**).

The three short α-helical segments in *Sm*Lip form a helical domain that is also present in the *Ec*TssJ structure (**Figure 3.11A**). Deletion of this domain in *Ec*TssJ results in reduced expression *in vivo*, with no expression observed *in vitro* (Felisberto-Rodrigues *et al*., 2011).

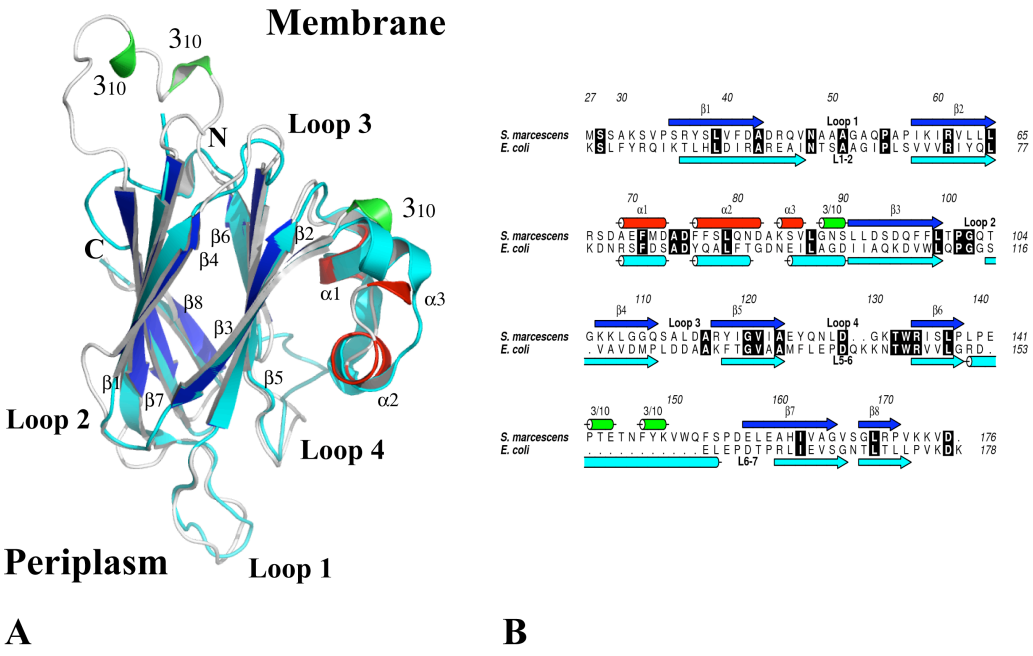


Figure 3.11 Comparison with the *Ec*TssJ crystal structure.
A) The structure of *Ec*TssJ (PDB 3RX9; cyan) is superimposed onto *Sm*Lip (coloured using the same colour scheme as **Figure 3.6A**). **B**) A structure-based sequence alignment of *Sm*Lip and *Ec*TssJ. The alignment is coloured as in **A**. Note that the signal peptide sequence has been excluded for clarity.

Analysis of the *Sm*Lip structure suggests the possibility that loop 1 may be involved in protein - protein interactions with other components of the T6SS (**Figure 3.11A**; Rao *et al*., 2011). This was confirmed in *Ec*TssJ, with the deletion of the L1-2 loop (loop 1 in *Sm*Lip) resulting in the absence of Hcp in culture supernatant (an indicator of a functional T6SS; Bingle *et al*., 2008; Felisberto-Rodrigues *et al*., 2011;

see section 1.2.1). In addition, it was established that *Ec*TssJ interacts with the protein TssM (Felisberto-Rodrigues *et al*., 2011), a large protein with an N-terminal cytoplasmic domain, three transmembrane helices, a periplasmic domain and has ATPase activity (**Figure 3.10**; Ma *et al*., 2012). *Ec*TssJ was found to interact with the C-terminus of the periplasmic domain of TssM with a K_d of 2-4 μM via the L1-2 loop. Despite this loop also being present in *Sm*Lip (loop 1), there is little sequence conservation between them (**Figure 3.11B**). It was speculated that variation in the length and composition of this loop might confer specificity between *Ec*TssJ and TssM homologues (Felisberto-Rodrigues *et al*., 2011). Indeed, loop 1 in the *Sm*Lip structure is comprised of 14 residues, more than is observed in the L1-2 loop in *Ec*TssJ and is a possible consequence of variation between bacterial species within the T6SS.

A possible chaperone-like role for *Ec*TssJ has been proposed, ensuring proper folding or stability of the C-terminus of the periplasmic domain of TssM (Felisberto-Rodrigues *et al*., 2011). A number of lipoproteins in other bacterial secretion systems carry out chaperone-like functions. In the T2SS, PulS is an outer membrane anchored lipoprotein responsible for the insertion of channel forming proteins into the outer membrane (Hardie *et al*., 1996). Furthermore, the lipoprotein InvH in the T3SS aids in the localisation and insertion of InvG to the outer membrane, forming the outer membrane translocation complex (Crago and Koronakis, 1998; Daefler and Russel, 1998).

The interaction between *Ec*TssJ and TssM is essential for a functional T6SS, making the disruption between these two components a possible drug target (Felisberto-Rodrigues *et al*., 2011). Potentially a small peptide or small molecule could out compete the L1-2 loop (loop 1 in *Sm*Lip) for binding to the complementary cavity of TssM (Felise *et al*., 2008; Shahian *et al*., 2009). This may represent a suitable target to pursue for the development of novel inhibitors in order to disrupt the assembly of the T6SS.

CHAPTER 4

RESULTS AND DISCUSSION: Rap1b

4.1 Aims

The aim of this project was to structurally characterise *S. marcescens* resistance associated protein 1b (Rap1b). This protein is encoded on the operon of the recently identified *S. marcescens* T6SS (Murdoch *et al*., 2011), along with a number of other small auxiliary proteins (see section 1.3.1). A structure of Rap1b could not be modelled, therefore the crystal structure was sought to allow comparisons with known protein structures and potentially give insight into the role this protein has within the T6SS.

4.2 Recombinant Rap1b expression and purification

Full length Rap1b consists of 151 residues, with the first 32 residues constituting the signal peptide sequence targeting the protein to the periplasm (Petersen *et al*., 2011). The cloning of a truncated form of the *S. marcescens Rap1b* gene (residues 28-151) from genomic DNA into the pET15b-TEV vector was carried out by Grant English (University of Dundee).

Recombinant Rap1b was produced in *E. coli* Rosetta-gami™ (DE3) cells (Novagen). Cultures were grown in LB media (Bertani, 1951) supplemented with 100 μg L^{-1} carbenicillin before induction with isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1 mM for 24 hrs as described in section 2.2.2. The protein was purified as described in section 2.3 with the His-tag still present. The protein eluted from a size exclusion column as a single species of approximate mass of 30 kDa, corresponding to a dimer. Samples were taken for SDS-PAGE analysis (**Figure 4.1**) and fractions containing the protein were pooled and concentrated to 13 mg mL^{-1}. A final yield of 3 mg L^{-1} of bacterial culture was obtained. The theoretical mass of one subunit is estimated as 14 kDa with a calculated pI of 6.3. The purified protein sample was stored at 4 °C until further use.

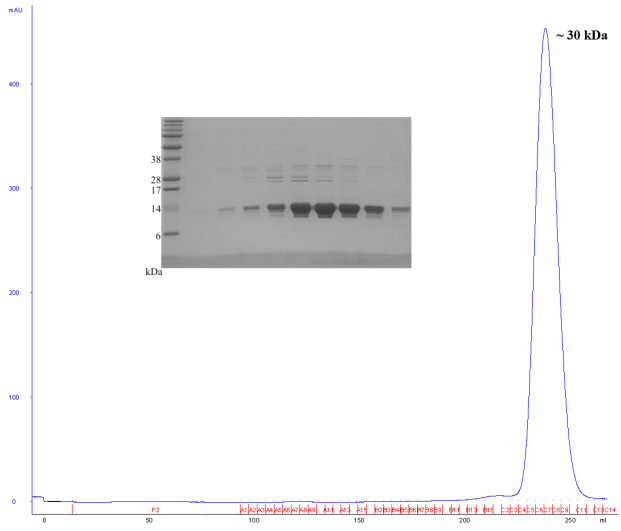


Figure 4.1 Size exclusion chromatography. A single absorbance peak at 280 nm was observed, corresponding to a protein of around 30 kDa. 4 mL fractions were collected.

4.3 Crystallisation and data processing

Initial crystallisation screens of Rap1b were carried out at 20 °C by the sitting drop vapour diffusion method in 96-well plates (see section 2.4.2). This was achieved with a Phoenix Liquid Handling System (Rigaku, Art Robins Instruments) and the commercially available PEG (Qiagen) and JCSG+ (Molecular Dimensions) screens. Crystallisation occurred at a protein concentration of 13 mg mL^{-1} and in the reservoir condition 0.5 M Li_2SO_4, 15 % PEG 8000 (**Figure 4.2**). This was further optimised using hanging drop vapor methods (see section 2.4.2) with drops consisting of 1 µL of protein at 8 mg mL^{-1} in 25 mM Tris-HCl, pH 7.5, 250 mM NaCl, 1 mM DTT and 1 µL of 0.55 M Li_2SO_4, 5 % PEG 8000. Optimised crystals grew to a maximum dimension of approximately 0.25 mm, over five days.

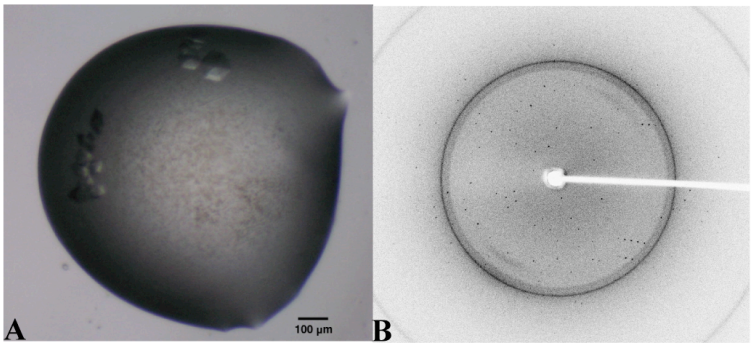


Figure 4.2 Rap1b crystals.
A) Initial crystals of Rap1b grew over three days. These crystals gave diffraction in-house to 2.4 Å (**B**).

A single crystal from the optimised condition was soaked for approximately 20 seconds in 125 mM aqueous sodium iodide. This was used to measure a 1.88 Å high-resolution dataset in-house using a Rigaku MicroMax-007 rotating-anode X-ray generator (copper K_α, λ=1.5418 Å) coupled to an R-AXIS IV^{++} image plate detector. The high concentration of Li_2SO_4 present in the crystallisation condition proved to be a suitable cryoprotectant. Data were indexed and integrated using *iMOSFLM* (Powell *et al*., 2007), with space group $P3_121$ and unit cell lengths $a=b=$ 77.9 Å and $c=$ 50.6 Å (**Table 4.1**). Data were scaled and the space group was suggested using *SCALA* and *POINTLESS* (Evans, 2006) from the *CCP4* program suite (Winn *et al*., 2011). A Matthews coefficient (Matthews, 1968) of 2.72 $Å^3$ Da^{-1} was calculated with 55 % solvent content, assuming a single polypeptide chain in the asymmetric unit.

4.4 Structure determination of Rap1b

4.4.1 Phasing

Sequence analysis of Rap1b using *BLAST* (Altschul *et al*., 1997) and the *PHYRE* protein fold recognition server (Kelley and Sternberg, 2009) indicated little or no homology to known proteins. The high-resolution diffraction data recorded in-house was used to solve the structure of Rap1b by SAD methods targeting potential iodides from the halide soak and endogenous sulfurs (Dauter *et al*., 2000; see section 2.5.3). Seven potential anomalous scattering positions were identified and experimental phases were calculated, producing a figure-of-merit of 0.28 to 1.88 Å resolution. Density modification was carried out using solvent flattening and histogram matching. Subsequent refinement confirmed two of the anomalous scattering positions as iodides (**Figure 4.3A**), and five were confirmed as well-ordered sulfurs present in the amino acid sequence, including two involved in disulfide bond formation (**Figure 4.3B**).

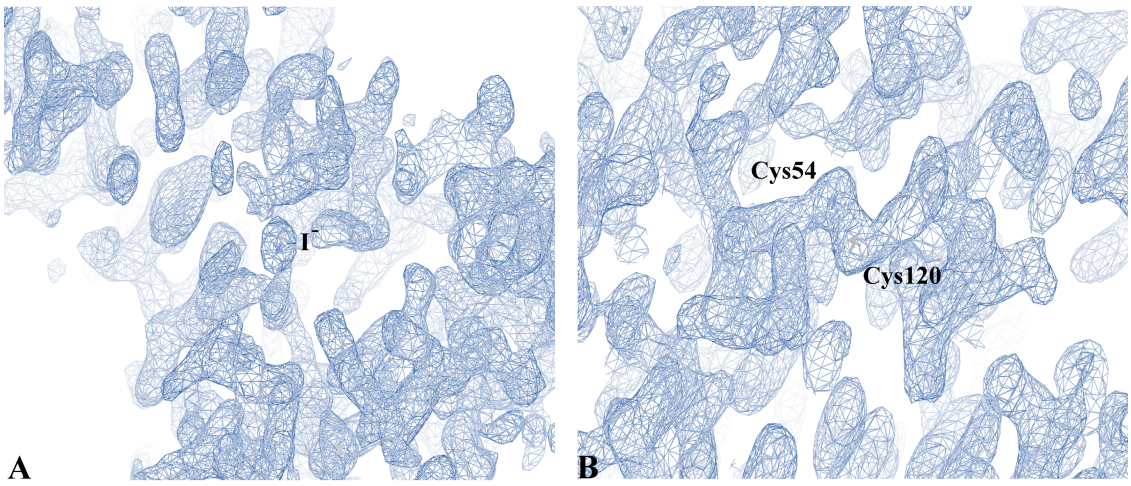


Figure 4.3 Density modified map.
Substructure searching identified both iodides (**A**) from the halide soak and well-ordered intrinsic sulfurs (**B**). Map contoured at 1.5 σ.

4.4.2 Model building and refinement

The initial model consisted of 108 residues, with a correlation coefficient of 0.51 and R_{work} and R_{free} values of 46.3 % and 52.9 % respectively. Subsequent model building in *COOT* (Emsley *et al*., 2010) extended this to 119 residues, with a correlation coefficient of 0.95 and R_{work} and R_{free} values improving to 18.3 % and 22.1 % respectively. Refinement calculations were performed using *REFMAC5* (Murshudov *et al*., 2011) and translation/libration/screw analysis (TLS) was applied (Painter and Merritt, 2006). The refinement proceeded with the incorporation of water molecules, iodides, ethylene glycol and a number of side chains with dual rotamer conformations. The refinement was terminated when there were no significant changes in R_{work} and R_{free} and inspection of the difference density map suggested that no further corrections or additions were justified. The geometry and quality of the model was validated using *MOLPROBITY* (Chen *et al*., 2010) and the subsequent Ramachandran plot reveals 100 % of all residues in most favoured or allowed regions. Crystallographic statistics are given in **Table 4.1**.

Table 4.1 Crystallographic statistics of Rap1b.

	Rap1b
Space group	$P3_121$
a, b, c (Å)	77.9, 77.9, 50.6
Resolution[a] (Å)	67.5 – 1.88 (1.98 – 1.88)
No. reflections recorded	162739 (21729)
Unique reflections	14813 (2130)
Completeness (%)	100.0
Multiplicity / $<I/\sigma I>$	11.0 (10.2) / 26.6 (5.9)
Anomalous completeness (%)	100.0
Anomalous redundancy	5.6 (5.1)
Wilson *B* ($Å^2$)	28.3
Residues / waters / ligands	119 / 117 / 7
R_{merge}[b] (%)	5.3 (38.3)
R_{work}[c], R_{free}[d] (%)	18.3 / 22.1
Ave. *B*-factor ($Å^2$)	
Protein	21.1
Waters, iodides, ethylene glycol	42.3, 49.4, 55.7
Cruickshank DPI[e] (Å)	0.1
Ramachandran plot	
Most favoured	118 residues
Additional allowed	1 residue
Outliers	0
RMSD on ideal values[f]	
Bond lengths (Å) / angles (°)	0.02 / 1.51

[a.] Values in parentheses refer to the highest resolution shell. [b.] $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - <I(hkl)>|/ \sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $<I(hkl)>$ is the mean value of $I_i(hkl)$ for all *i* measurements. [c.] $R_{work} = \sum_{hkl}||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. [d.] R_{free} is the same as R_{cryst} except calculated with a subset, 5 %, of data that are excluded from the refinement calculations. [e.] Diffraction Precision Index (Cruickshank, 1999). [f.] (Engh and Huber, 1991).

4.5 The Rap1b structure

4.5.1 Overall structure

Rap1b displays a compact globular structure with five α-helices (α2, α3, α4, α5 and α6) and two short α-helical segments (α1 and α7) located at either termini (**Figure 4.4A**). The large loop between helices α4 and α5 corresponds to a chain break in the crystal structure, where it was not possible to interpret diffuse and weak electron density. Consequently residues Glu108 to Asp111 are absent from the model. A single

molecule of Rap1b is present in the asymmetric unit, however analysis of interactions using *COOT* (Emsley *et al*., 2010) indicates this can be superimposed on to a second molecule by the symmetry operator -x,-x+y,-z+1/3 to form a dimer (**Figure 4.4A**). The ASA of one molecule is approximately 8250 $Å^2$, while the interface between the two molecules covers an area that is approximately 24 % of the ASA (Krissinel and Henrick, 2007).

During the course of these studies, the crystal structure of another of the small proteins, Rap2b, became available from work carried out by Dr. V. Srikannathasan (University of Dundee). Rap2b (**Figure 4.4B**) displays a compact globular structure like Rap1b, however Rap1b is longer with additional residues at the N-terminus and the C-terminus forming an additional short α-helix (α7; **Figure 4.4C**). Rap2b has four molecules in the asymmetric unit forming two dimers by the subunits A:B and C:D. The four molecules are comparable, with an RMSD between superimposed C_α atoms ranging from 0.5 Å (monomer B and D) to 0.7 Å (monomers A and C) with an average value of 0.58 Å. The ASA averages out at approximately 6330 $Å^2$; the range is from 6240 $Å^2$ for molecule B to 6470 $Å^2$ for molecule D. Each dimer covers an area that is approximately 20 % of the ASA. The large ASA used for dimer formation in both Rap1b and Rap2b suggests a biological relevance (Krissinel and Henrick, 2007) of these stable oligomers and is consistent with both proteins existing as stable dimers in solution (data not shown).

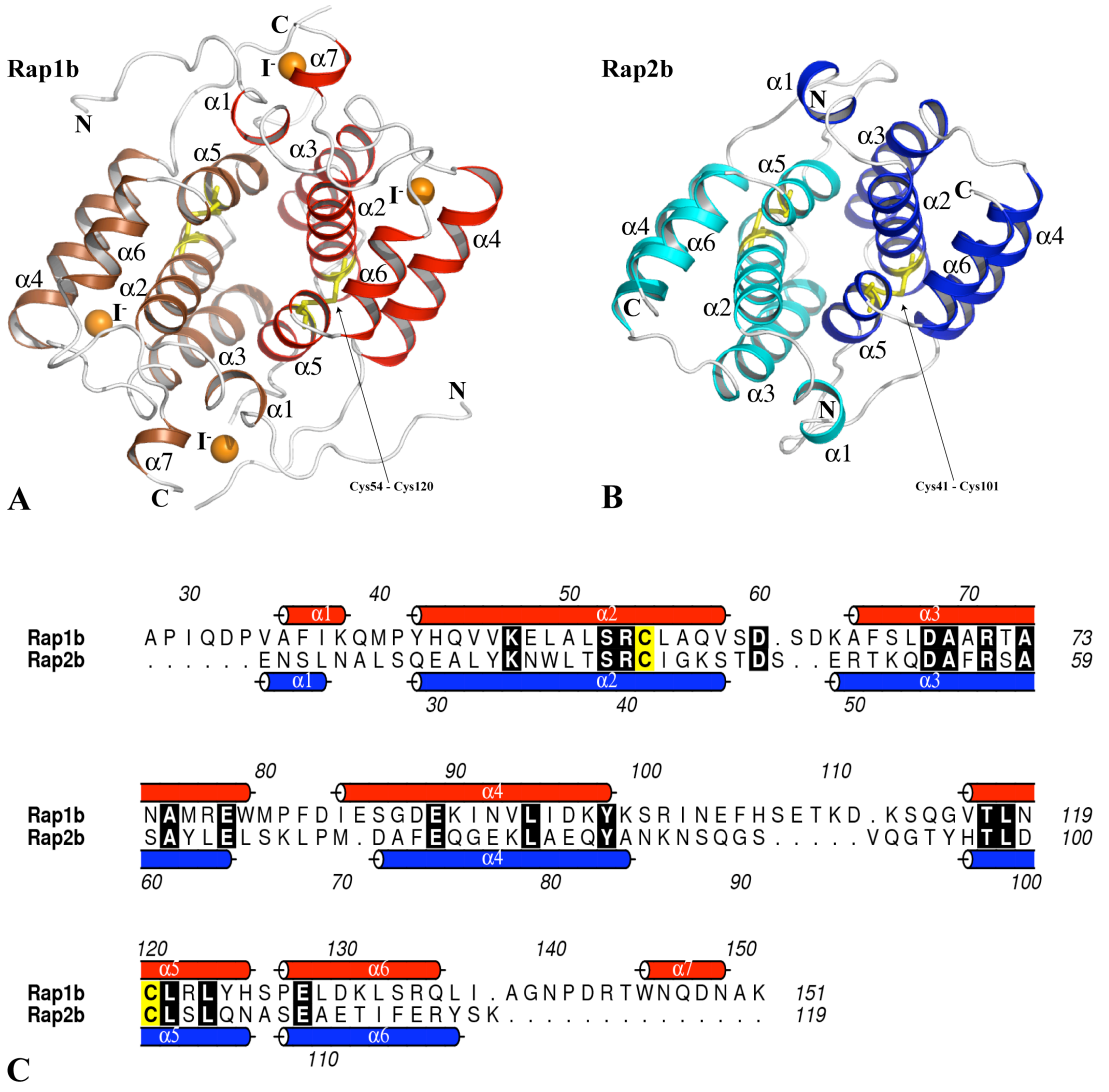


Figure 4.4 Rap1b and Rap2b comparison.
A) The structure of Rap1b is coloured with α-helices red. The α-helices in the symmetry related molecule are coloured brown. Iodides used in phasing are coloured orange. **B**) The structure of Rap2b is coloured blue, with the second molecule in the asymmetric unit coloured cyan. **C**) A structure-based sequence alignment highlights the conserved secondary-structure content in Rap1b (red cylinders) and Rap2b (blue cylinders). Residues involved in disulfide bond formation are coloured yellow. The signal peptide has been excluded.

Despite a low sequence identity of approximately 20 % (**Figure 4.4C**), both Rap proteins display an obvious structural resemblance with one another (**Figure 4.5**). Superimposing the two structures matches 71 residues with an RMSD of 1.4 Å (**Figure 4.5**). A conserved disulfide bond (Cys54 – Cys120 in Rap1b and Cys41 – Cys101 in Rap2b) brings together helices α2 and α5 and helps form a hydrophobic core. This core is comprised of helices α2, α3 and α5 interacting with the corresponding helices from

the second molecule across the two-fold axis of symmetry (**Figure 4.4A & B**). This interaction contributes to dimer stabilisation by virtue of extensive van der Waals and hydrogen bond formation.

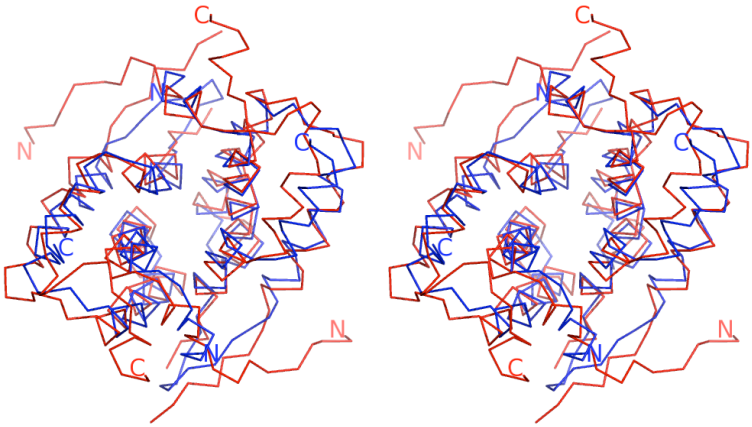


Figure 4.5 A C_α trace of Rap1b and Rap2b.
Stereoview overlay of Rap1b (red) superimposed on to Rap2b (blue). Both structures are orientated as in **Figure 4.4**.

To investigate the possible function of both Rap1b and Rap2b, a search for structural homologues using *PDBeFold* (http://pdbe.org/fold) was carried out. Using the monomer and crystallographic dimer of Rap1b as search models, the results indicated no similarity with other protein structures deposited in the PDB. The closest match gave a Z-score of 2.7, suggesting low statistical significance. Furthermore, a search using both the monomer and dimer of Rap2b as search models gave a Z-score of 2.2. Taken together, these results suggest the overall fold observed in both Rap1b and Rap2b is previously uncharacterised.

4.5.2 Disulfide bond formation

Several anomalous scattering positions were identified during SAD phasing and correspond to well-ordered sulfur atoms present in the amino acid sequence of Rap1b (**Figure 4.4C**). In particular, two of the identified sulfur positions correlate to cysteine residues involved in disulfide bond formation; Cys54 on the α2-helix and Cys120 on

the α5-helix (**Figure 4.4A & C**). In addition, the structure of Rap2b also has a single disulfide bond present, comprising residues Cys41 and Cys101 (**Figure 4.4B & C**). An overlay of Rap1b onto Rap2b aligns the disulfide positions in both structures (**Figure 4.6**). This disulfide bond formation takes place in the periplasm and is likely introduced by the thiol-disulfide oxidoreductase, DsbA (Shouldice *et al*., 2011). During initial expression tests of recombinant Rap1b and Rap2b, it was noted that soluble protein was produced only in *E. coli* Rosetta-gami™ (DE3) cells (Novagen); a cell line that promotes the formation of disulfide bonds (data not shown). This suggests the covalent bond is necessary for correct folding to occur and for enhanced stability of the proteins. The stabilisation effect of introducing a disulfide bond into a protein has been studied previously in T4 lysozyme (Matsumura *et al*., 1989; Wetzel *et al*., 1988). Furthermore, the presence of a disulfide bond may infer resistance to proteolytic degradation in the periplasm (Beck *et al*., 2008; Ireland *et al*., 2006).

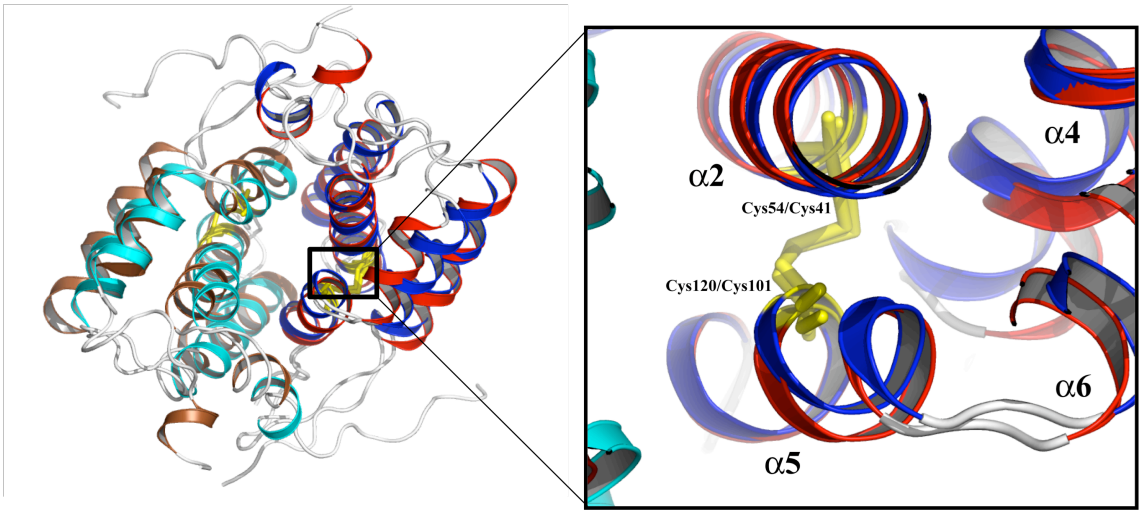


Figure 4.6 Disulfide bond.
The position of the disulfide in both Rap1b and Rap2b align with one another. The same orientation and colour as **Figure 4.4** is used.

4.6 Discussion

An active T6SS was recently identified in the opportunistic Gram-negative pathogen *S. marcescens* (Murdoch *et al*., 2011). A cluster of genes located in the middle of the T6SS operon encodes several small auxiliary proteins (**Figure 4.7A**; English *et al.*, 2012; Murdoch *et al*., 2011). These small proteins can be grouped based on their pI, secondary structure prediction and the presence or absence of a signal peptide sequence (**Figure 4.7A**; see section 1.3.1).

The proteins Rap1b and Rap2b are members of the Rap family of self-resistance proteins that have recently been characterised in the *S. marcescens* T6SS (**Figure 4.7A**; English *et al.*, 2012). This family of proteins has been shown to confer protection against the secreted toxins Ssp1 and Ssp2 (**Figure 4.7B**; English *et al.*, 2012) and is analogous to the Tsi and Tse substrates of the *P. aeruginosa* HSI-1 T6SS (Hood *et al.,* 2010; Russell *et al.,* 2011; see section 1.3). The amino acid sequence of Rap proteins shows the presence of a signal peptide locating them to the oxidative periplasmic space (**Figure 4.7B**; Peterson *et al.*, 2011) for correct folding and disulfide formation to occur (Shouldice *et al*., 2011). Homologues have been identified in several bacteria with members displaying a conserved disulfide bond and predicted helical fold (**Figure 4.7C**; English *et al.,* 2012).

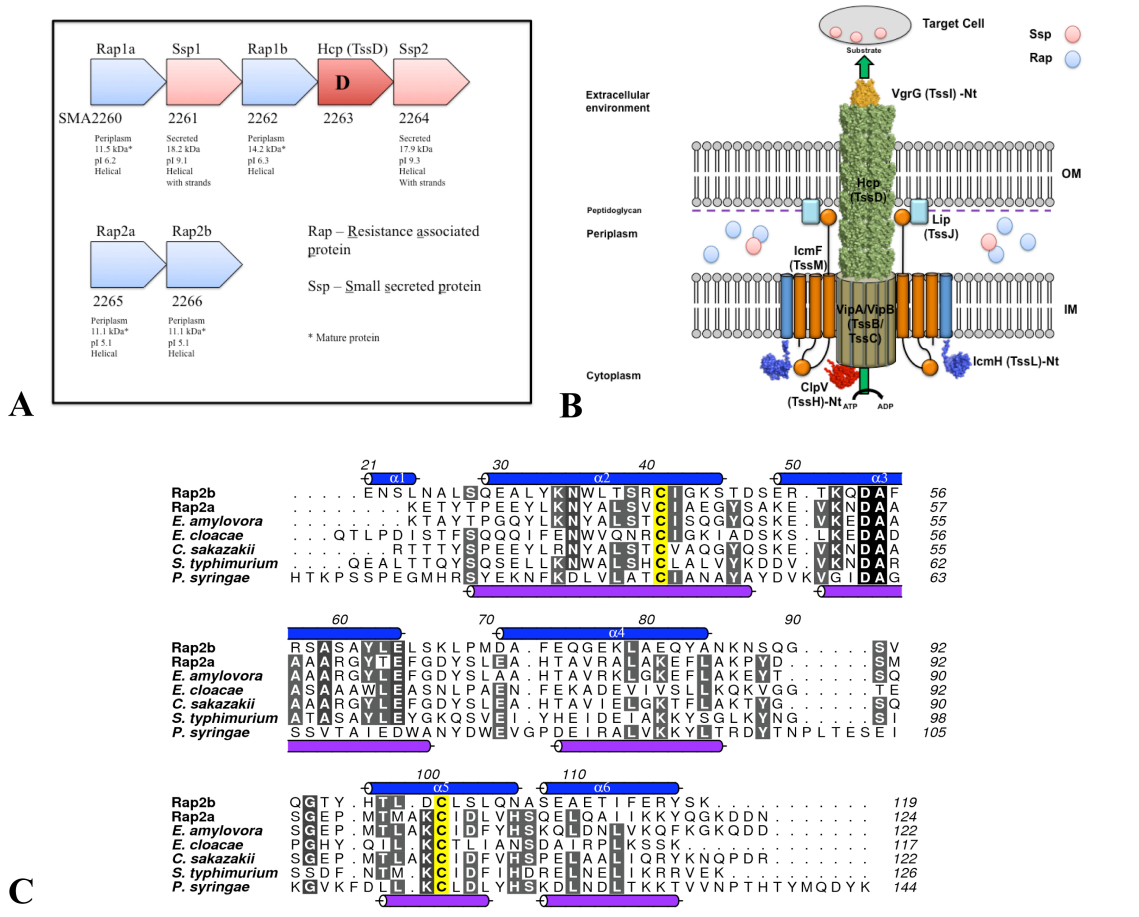


Figure 4.7 Toxins and self-resistance proteins.
A) These small proteins have been identified as secreted toxins (Ssp) and cognate self-resistance proteins (Rap). Figure adapted from (English *et al*., 2012). **B**) The secreted proteins (Ssp) may be delivered to the periplasm of target cells as with Tse1 and Tse3 (Russell *et al*., 2011). To prevent self-intoxication, self-resistance (Rap) proteins may be localised to the periplasm of the producing cell, binding to and negating the affects of the Ssp proteins. Figure adapted from (Bönemann *et al*., 2010; Cascales and Cambillau, 2012). **C**) Sequence alignment of *S. marcescens* Rap2b and Rap2a with homologous proteins from *Erwinia amylovora* (GenBank CBA22869.1), *Enterobacter cloacae* (NCBI Reference Sequence: YP_003612051.1), *Cronobacter sakazakii* (NCBI Reference Sequence: YP_001439955.1), *Salmonella typhimurium* (NCBI Reference Sequence: NP_459276.1) and *Pseudomonas syringae* (NCBI Reference Sequence: YP_237109.1). The secondary structure of Rap2b (blue cylinders) and the predicted secondary structure of the *P. syringae* protein (purple cylinders) are shown. Cysteine residues involved in disulfide bond formation are colored yellow. The signal peptide sequences have been excluded.

The structures of Rap1b and Rap2b reveal an α-helical fold that is previously uncharacterised and is representative of Rap proteins (**Figure 4.7C**). Currently, the only other structure of an immunity protein associated with the T6SS is the cytoplasmic Tsi2 protein from the HSI-1 T6SS of *P. aeruginosa* (**Figure 4.8A**; PDB 3RQ9 & 3STQ; Li *et al*., 2012; Zou *et al*., 2012). Tsi2 is an acidic protein that exhibits a novel helical fold and forms a dimer in solution, features that are shared with Rap proteins. However the structure of Tsi2 and the Rap proteins are unrelated (**Figure 4.8B**). Superimposing

Rap1b onto Tsi2 matches few residues, with the majority of secondary structure elements being distinct (**Figure 4.8B**).

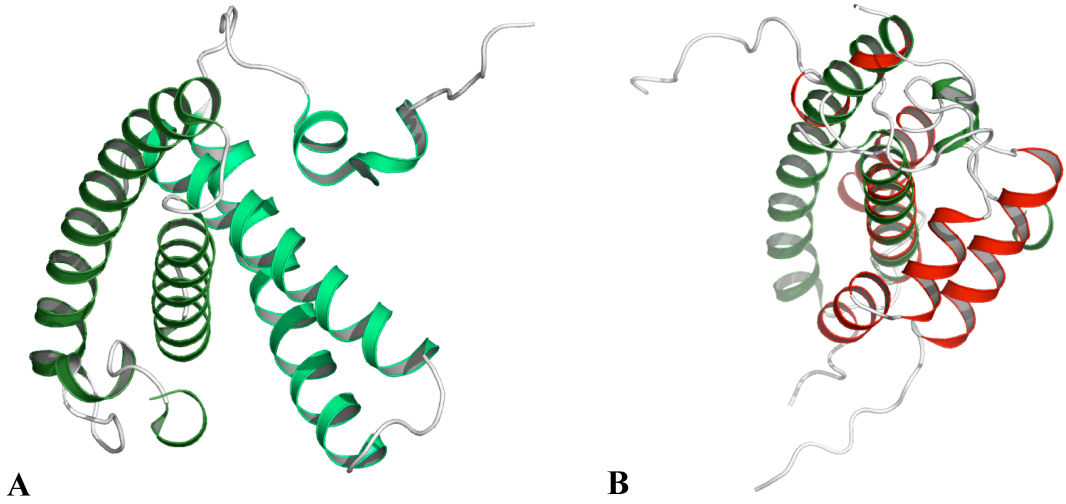


Figure 4.8 Tsi2 structure.
A) The structure of the Tsi2 dimer (PDB 3RQ9; Li *et al*., 2012). **B**) A single molecule of Tsi2 (green) superimposed onto a single molecule of Rap1b (red). For clarity the symmetry related Rap1b molecule is excluded. The same orientation as **Figure 4.4** is used.

A common feature of toxin-antitoxin pairs is the opposing electrostatic properties (Yamaguchi and Inouye, 2011). This is observed with the secreted Ssp toxins (pI 9) and cognate Rap resistance proteins (pI 5-6; **Figure 4.7A**). Interestingly, the analysis of the Rap1b and Rap2b structures reveals a distinct crescent shaped basic patch on either side of the dimer (**Figure 4.9**). Residues Lys38, Lys90, Lys97, Arg101, Lys131, Arg143 and Lys151 (residues Lys67, Lys78, Lys86, Arg116 and Lys119 in Rap2b) form this patch (**Figure 4.9**). This basic crescent may help to orientate and ensure the correct binding to respective binding partners. The structure of a Ssp:Rap complex would provide further insight into the details of this interaction.

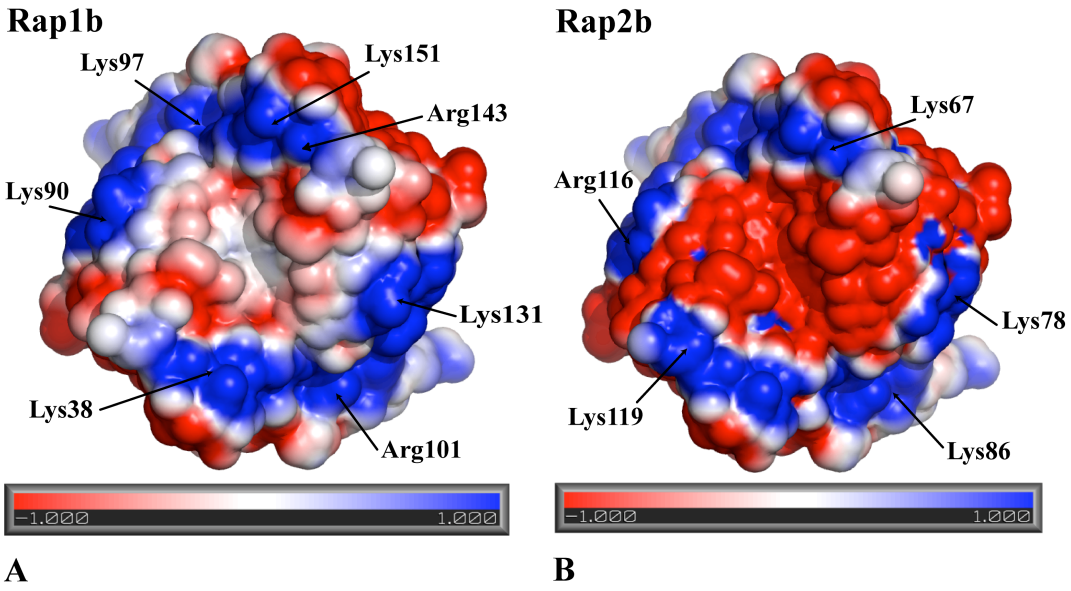


Figure 4.9 Electrostatic surface potential.
Calculated using the *APBS* plugin for PyMOL (Baker *et al*., 2001; Lerner, 2006; Schrodinger, 2010), the electrostatic charge is contoured at 1 kT/e and -1 kT/e and highlights the charge distrubution on Rap1b (**A**) and Rap2b (**B**). Negative (acidic) charge is red, positive (basic) charge is blue. Residues that contribute to the basic patches are identified.

While highly specific interactions have been observed and characterised for the proteins Ssp1:Rap1a and Ssp2:Rap2a (English *et al*., 2012), it is not clear what resistance the proteins Rap1b and Rap2b provide against. These proteins are formed in the cytoplasm and are transported to the periplasm via the Sec pathway in an unfolded state. Upon transport to the periplasm, formation of the conserved disulfide bond occurs to ensure the correct folding of the proteins, allowing the proteins to correctly bind and protect against their respective toxins. It may be possible that these proteins self-protect against other uncharacterised toxins, however this is unlikely as a common pattern observed is Rap resistance proteins located adjacent to their toxin partner (**Figure 4.7A**). A more likely role proposed for these proteins is to provide protection against similar toxins from different bacterial species, providing *S. marcescens* Db10 with a competitive advantage (English *et al*., 2012).

PART II

ASSESSMENT OF POTENTIAL GRAM-VE DRUG TARGETS

CHAPTER 5

INTRODUCTION: Fragment screening

5.1 The need for new antimicrobials

Hospital-acquired infections (HAIs) pose a major clinical challenge to public health, affecting more than 2 million patients in Europe and resulting in approximately 175,000 deaths annually (Chopra *et al*., 2008). In addition to the human cost of HAIs, the burden on healthcare budgets is estimated to be as high as $45 billion annually in the U.S. (Scott, 2009). As populations in developed countries get older and the number of immunosupressed individuals increases, this healthcare cost is set to rise (Chopra *et al*., 2008).

HAIs are typically associated with surgical procedures or invasive medical devices (Peleg and Hooper, 2010) and can cause urinary tract infections as well as lower respiratory and bloodstream infections (Gaynes *et al*., 2005; Peleg and Hooper, 2010). The Gram-negative bacteria predominantly responsible for HAIs are the *Acintobacter* species, *E. coli*, *Enterobacter* species, *Klebsiella pneumoniae*, *S. marcescens* and *P. aeruginosa* (**Table 5.1**; Gaynes *et al*., 2005; Hidron *et al*., 2008). In particular, the opportunistic pathogen *P. aeruginosa* is recognised as a significant cause of HAIs (Kerr and Snelling, 2009), accounting for around 18 % of all instances of hospital-acquired pneumonia (**Table 5.1**; Gaynes *et al*., 2005). Furthermore, antibiotic resistance in *P. aeruginosa* and other bacteria is reported to be increasing (Hidron *et al*., 2008; Obritsch *et al*., 2004; Souli *et al*., 2008), resulting in a significant impact on the treatment of HAIs (Arias and Murray, 2009; Calfee, 2012). Instances of multidrug resistant bacteria, where resistance to three or more classes of antibiotics is observed (Souli *et al*., 2008), is reported to be 16 % of all HAIs in the U.S. during the period 2006-2007 (Hidron *et al*., 2008).

Table 5.1 Bacterial isolates identified from U.S. intensive care units in 2003.

	Percentage of isolates, by infection type			
Gram-negative pathogen	**Pneumonia (*n* =4365)**	**Bloodstream infection (*n* = 2351)**	**Surgical site infection (*n* = 2984)**	**Urinary tract infection (*n* = 4109)**
Acintobacter species	6.9	2.4	2.1	1.6
E. coli	5.0	3.3	6.5	26.0
Enterobacter species	10.0	4.4	9.0	6.9
K. pneumoniae	7.2	4.2	3.0	9.8
S. marcescens	4.7	2.3	2.0	1.6
P. aeruginosa	18.1	3.4	9.5	16.3
Other	14.1	3.8	9.8	10.7

Data taken from (Gaynes *et al*., 2005).

5.1.1 Antimicrobial drug discovery pipeline

Many large pharmaceutical companies have abandoned or scaled back their antimicrobial research, with few drugs currently in preclinical development (Boucher *et al*., 2009; Talbot *et al*., 2006). The stringent regulatory environment worldwide and concerns regarding length of patents (Fernandes, 2006; Katz *et al*., 2006) have contributed to the increase in costs associated with drug discovery (Bunnage, 2011; Paul *et al*., 2010). Consequently, pharmaceutical companies are now ensuring their investments generate substantial revenues to offset research and development costs (Vicente *et al*., 2006). This has resulted in a shift in focus towards the research and development of new drugs for treating chronic diseases, as revenue is maintained over a prolonged period of time in comparison to bacterial infections (Vicente *et al*., 2006). This lack of novel drugs in the drug discovery pipeline (Boucher *et al*., 2009; Talbot *et al*., 2006), coupled with the increase in antibiotic resistance in bacteria (Hidron *et al*.,

2008; Souli *et al*., 2008), has the potential to cause future problems for the treatment of infectious diseases.

5.2 Target assessment

On average, a single drug takes approximately 13 years and close to $1 billion to get to the clinic from the drug discovery phase (**Figure 5.1**; Bunnage, 2011; Paul *et al*., 2010). This makes selecting potential drug targets an important step in drug discovery, as failure to select an appropriate target will be costly and resources will have been expended needlessly (**Figure 5.1**; Bunnage, 2011). The availability of complete genome sequences (http://www.ncbi.nlm.nih.gov/genome) has made the assessment and prioritisation of targets attainable. Databases such as the AEROPATH Target Database (http://aeropath.lifesci.dundee.ac.uk/) aid this process by amalgamating data from a number of sources and providing a pool of putative targets to assess.

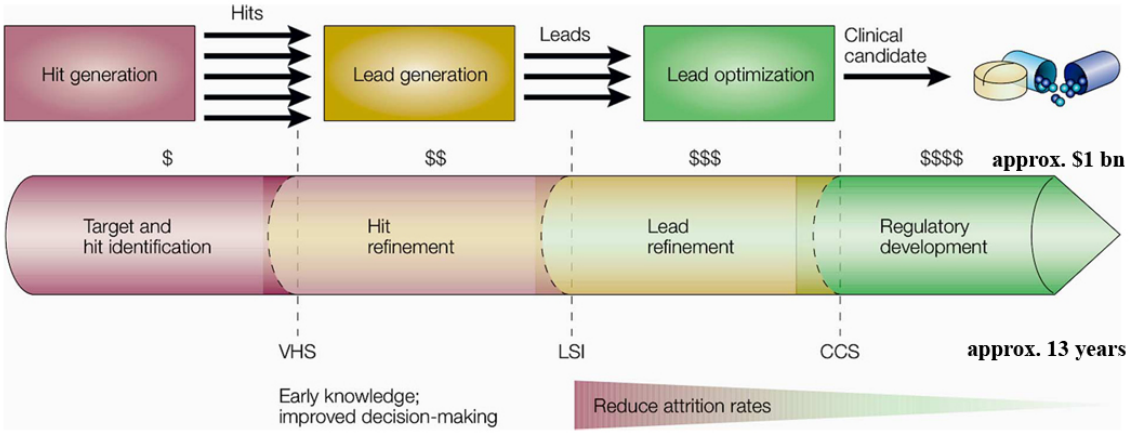


Figure 5.1 Drug discovery process.
The stages of the drug discovery process are illustrated here, highlighting the increase in costs as a molecule progresses. The attrition rate of potential drugs at each stage results in few drugs reaching the market. VHS; validated hit series, LSI; lead series identified, CCS; clinical candidate selection. Figure edited from (Warrick *et al*., 2008).

Each putative drug target is assessed against the criteria highlighted in **Table 5.2** and scored using a traffic light system to prioritise them (Frearson *et al*., 2007). Favourable scoring is indicative of a good quality target to pursue and therefore

underpins the entire drug discovery process (Bunnage, 2011; Frearson *et al*., 2007). A key criterion is the concept of druggability (Hopkins and Groom, 2002); is the target amenable to functional manipulation through interactions with small-molecule drugs that adhere to the Lipinski rule of five (Lipinski *et al*., 2001)? Predicting druggability (Cheng *et al*., 2007; Fauman *et al*., 2011) can streamline and enhance the process of target validation, ensuring targets most likely to succeed in the hit and lead identification process are selected (Hajduk *et al*., 2005).

Table 5.2 Traffic light definitions for target assessment.

Criterion	Red	Amber	Green
Target validation	No or weak evidence that the target is essential for growth or survival.	Either genetic or chemical evidence that target is essential for growth or survival.	Genetic and chemical evidence that target is essential for growth or survival.
Druggability	No drug-like inhibitors are known and active site of target is not druggable.	Drug-like inhibitors are known or active site potentially druggable.	Drug-like inhibitors are known and druggable active site (i.e. clinical precedent within the target family.)
Assay feasibility	No *in vitro* assay developed and / or significant problems with reagents (cost or supply).	*In vitro* assay exists, development into plate format feasible, but not achieved.	Assay ready in plate format and protein supply assured within appropriate timelines.
Toxicity	Human homologue present and little or no structural or chemical evidence that selective inhibition is possible.	Human homologue present but some structural or chemical evidence that selective inhibition is possible.	No human homologue present or human homologue known to be non-essential.
Resistance potential	Target has multiple gene copies or isoforms within the same species and is subject to escape from inhibition.	Target has isoforms within the same species or might be subject to escape from inhibition.	Target has no known isoforms within same species and is not subject to escape from inhibition.
Structural information	No structure of target or closely related homologue.	Structure without ligand available and / or poor resolution (> 2.3 Å) or opportunity to build a good homology model.	Ligand bound structure of target (or ligand in closely related homologue) available at high resolution (< 2.3 Å).

(Frearson *et al*., 2007).

5.3 Fragments

The pharmaceutical industry uses high-throughput screening (HTS) of large chemical libraries to identify lead compounds that are developed to become marketed drugs (Erlanson *et al*., 2004; Hartshorn *et al*., 2005; Rees *et al*., 2004). However, over 90 % of compounds in preclinical development fail to progress to clinical trials as a result of poor target validation or failure in lead optimisation (**Figure 5.1**), resulting in compounds that are either toxic or have poor drug-like properties (Brown and Superti-Furga, 2003). The 'rule of five' (Lipinski *et al*., 2001) was proposed as a framework to predetermine drug-like compounds most likely to have poor solubility/permeability. Lipinski and colleagues observed that this is more likely to occur in compounds that have; a molecular weight greater than 500 Da, *c*log*P* (lipophilicity) greater than 5, more than 5 hydrogen bond donors and 10 hydrogen bond acceptors (Lipinski *et al*., 2001). This framework can be used as a filter that can be applied to large HTS or virtual chemical libraries to aid in identifying lead compounds.

The design and selection of chemical libraries can contribute to successful lead discovery and subsequent optimisation (Leach *et al*., 2006). This has resulted in a shift towards screening libraries comprised of smaller molecules and has been termed fragment screening (Erlanson *et al*., 2004; Leach *et al*., 2006; Scott *et al*., 2012). Fragments are small, weak binding molecules that can be developed and optimised into lead compounds with greater binding affinity and specificity (**Figure 5.2**; Ciulli and Abell, 2007; Leach *et al*., 2006; Scott *et al*., 2012). The premise for starting with such small molecules in the drug discovery process was the observation that in improving properties such as potency, the complexity and molecular weight of lead compounds also increases (**Table 5.3**; Hann *et al*., 2001).

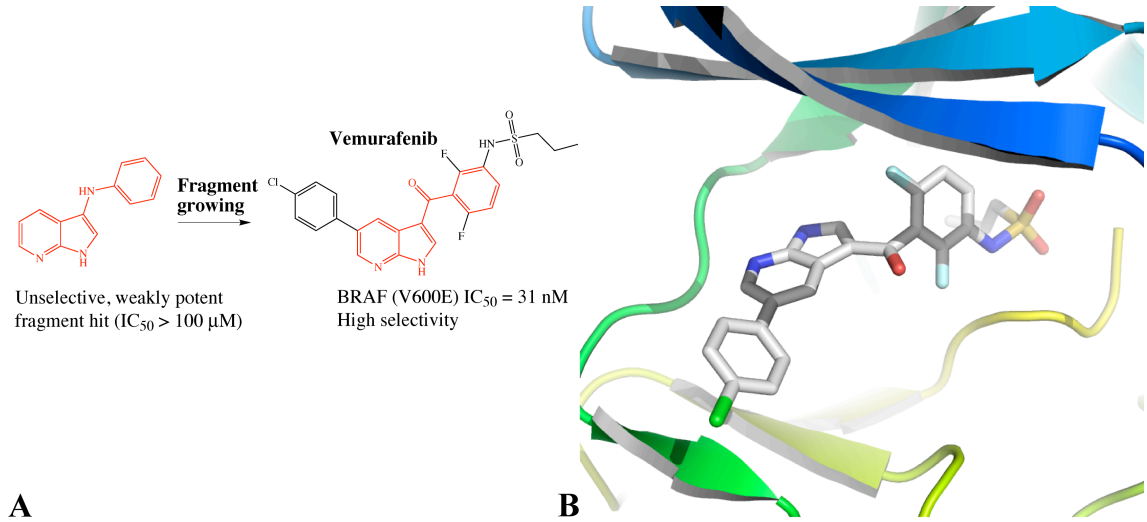


Figure 5.2 An example of fragment-based drug discovery.
A) A fragment was identified (red) and optimised to produce the B-Raf inhibitor Vemurafenib **B**) The crystal structure of B-Raf with Vemurafenib bound (PDB 3OG7; Bollag *et al*., 2010). Figure adapted from (Hoelder *et al*., 2012).

Table 5.3 Change in properties during optimisation from lead to drug.

Property	Average value for leads	Average value for drugs
Mw	272.0	314.0
H-bond donors	0.8	0.8
H-bond acceptors	2.2	2.5
*C*log*P*	1.9	2.4
Number of heavy atoms	19	22

Data taken from (Hann *et al*., 2001).

Thus in order to satisfy the Lipinski rule of five (Lipinski *et al*., 2001) and have favourable pharmacokinetic properties, starting molecules must be smaller to accommodate the optimisation process. The low molecular weight and complexity of fragments means they are highly efficient ligands (Hopkins *et al*., 2004) that can sample chemical space effectively (Lipinski and Hopkins, 2004). The simplistic nature of fragments results in few interactions with biological targets, giving rise to low binding affinities (Leach *et al*., 2006). Therefore detecting such weak binding requires a number

of biophysical techniques including bio-layer interferometry (BLI; Wartchow *et al*., 2011), differential scanning fluorimetry (DSF; Pantoliano *et al*., 2001), nuclear magnetic resonance (NMR; Pellecchia *et al*., 2008), surface plasmon resonance (SPR; Hoa *et al*., 2007) and X-ray crystallography (Hartshorn *et al*., 2005; Murray and Blundell, 2010). Libraries used for fragment screening typically comprise around 1,000 fragments and have been developed to adhere to a 'rule of three' (Congreve *et al*., 2003) framework that is akin to the Lipinski rule of five (Lipinski *et al*., 2001).

5.4 Aims

Several potential drug targets from *Acinetobacter baumannii*, *B. pseudomallei* and *P. aeruginosa* (**Table 5.4**) were selected for assessment based on the criteria described in **Table 5.2** (Frearson *et al*., 2007). The search for fragments that have the potential to generate lead compounds for further studies has been carried out (**Table 5.4**) using several screening techniques described in section 2.7.

Table 5.4 Potential drug targets screened

Target	Function	Screened using:
*Ab*Accc/*Pa*AccC	Catalyses the first step in fatty acid biosynthesis	DSF
*Pa*IMPDH	Catalyses the rate-limiting step in the biosynthesis of guanine nucleotides	DSF/BLI
*Pa*IspF	Involved in isoprenoid biosynthesis	BLI
*Ab*LpxC/*Pa*LpxC	Catalyses the first committed step in the biosynthesis of Lipid A	DSF
*Pa*MurA	Catalyses the first step in the biosynthesis of peptidoglycan	DSF
*Bp*PBP3/*Pa*PBP3	Catalyses the cross-linking of peptidoglycan	DSF/BLI/SPR
*Ab*RibD	Involved in riboflavin biosynthesis	BLI

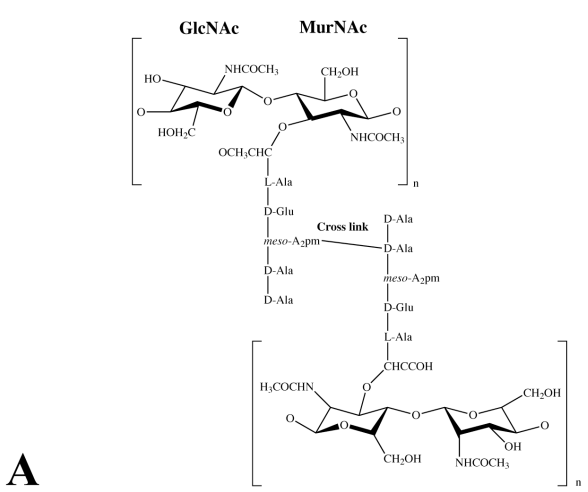
From these potential drug targets, two were selected for further characterisation (**Table 5.4**) and are the focus of the research described in chapters six and seven. The successful structure determination of both drug targets in complex with a ligand would aid in the understanding of the interactions required for specificity and selectivity and underpin the assessment of any identified chemical fragments.

CHAPTER 6

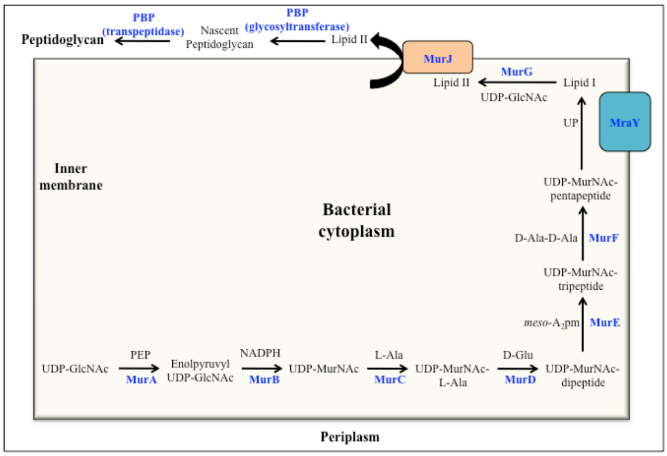
RESULTS AND DISCUSSION: PENICILLIN-BINDING PROTEIN 3

6.1 Background

Peptidoglycan is a macromolecule that forms a continuous layer on the exterior of the cytoplasmic membrane of bacteria (van Heijenoort and Gutmann, 2000; Vollmer *et al*., 2008). It is formed by β-1,4 linked *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) repeating units that are cross-linked via peptide bridges (**Figure 6.1A**; van Heijenoort, 2001; Vollmer and Bertsche, 2008; Vollmer *et al*., 2008). This cross-linking occurs at the final stages of peptidoglycan biosynthesis (**Figure 6.1B**) and is carried out by membrane-bound penicillin-binding proteins (PBPs; Lovering *et al*., 2012; Vollmer *et al*., 2008).



A



B

Figure 6.1 Peptidoglycan biosynthesis. **A**) The repeating unit of peptidoglycan in Gram-negative bacteria. **B**) The biosynthesis of peptidoglycan is a complex process comprising several enzymes (blue). Penicillin-binding proteins (PBPs) catalyse the polymerisation of Lipid II to nascent peptidoglycan strands and the cross-linking of strands to form mature peptidoglycan. Image adapted from (Lovering *et al*., 2012).

Peptidoglycan provides structural rigidity and the ability to withstand high internal osmotic pressures that would otherwise rupture the bacterial cell (van Heijenoort and Gutmann, 2000; Vollmer and Bertsche, 2008). The disruption or inhibition of peptidoglycan biosynthesis therefore results in cell lysis, making this pathway an ideal drug target (Kluge and Petter, 2010; Schneider and Sahl, 2010). This is demonstrated by antibiotics that target different stages of peptidoglycan biosynthesis (Lovering *et al*., 2012; Schneider and Sahl, 2010; Spratt and Cromie, 1988).

6.2 Aims

The aims of this project were to assess PBP3 from both *P. aeruginosa* (*Pa*PBP3) and *B. pseudomallei* (*Bp*PBP3) by producing soluble protein and identifying chemical fragments that bind (Ciulli and Abell, 2007; Leach *et al*., 2006; Scott *et al*., 2012). This information would inform as to whether or not it is possible to find and develop novel inhibitors. The crystal structure of PBP3 was sought to complement the fragment screening process by validating potential hits.

6.3 Recombinant *Pa*PBP3 & *Bp*PBP3 expression and purification

A truncated form of the *P. aeruginosa ftsI* gene encoding the amino acid residues 35 – 579 was amplified from genomic DNA and cloned into the pOPINF vector containing a non-cleavable N-terminal His-tag. This construct was provided by the Oxford Protein Production Facility (OPPF; Sainsbury *et al*., 2011). In addition, the truncated form of the *B. pseudomallei ftsI* gene encoding residues 65 – 584 was purchased from GenScript™ with *Nde*I/*Xho*I restriction sites and cloned into the pET15b-TEV vector (see section 2.2.1). The *B. pseudomallei ftsI* gene was also cloned into a modified version of this vector incorporating a N-terminal BAP sequence (Martin Zoltner, University of Dundee).

Recombinant *Pa*PBP3 and *Bp*PBP3 were produced in *E. coli* BL21 (DE3) pLysS (Stratagene) and Gold (Novagen) cells respectively. Cultures were grown in AI media (Studier, 2005) supplemented with 50 μg L^{-1} carbenicillin as described in section 2.2.2. The protein was purified and eluted from a size exclusion column (see section 2.3) as a single species of approximate mass of 60 kDa, corresponding to a monomer. Samples were taken for SDS-PAGE analysis (**Figure 6.2**) and fractions containing the protein were pooled and concentrated to approximately 10 mg mL^{-1}. A final yield of 2 mg L^{-1} (*Pa*PBP3) and 16 mg L^{-1} (*Bp*PBP3) of bacterial culture were obtained. The purified protein was stored at 4 °C until further use.

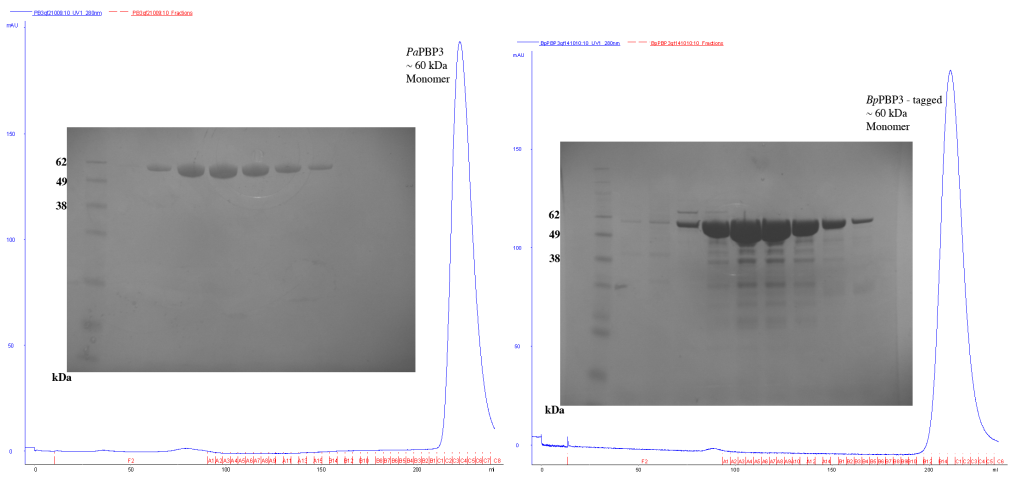


Figure 6.2 Size exclusion chromatography.
A single absorbance peak at 280 nm corresponds to a monomer of around 60 kDa for both *Pa*PBP3 (left) and *Bp*PBP3 (right). 4 ml fractions were collected.

6.4 Fragment screening

6.4.1 DSF

DSF (Niesen *et al*., 2007; Pantoliano *et al*., 2001; see section 2.7.3) was used first to demonstrate that recombinant *Pa*PBP3 and *Bp*PBP3 bind β-lactam antibiotics and to determine their effect on protein stability. The thermal profile of *Pa*PBP3 and *Bp*PBP3 display single melting curves with a melting temperature (T_m) of 37.3 ºC and 48.7 ºC

respectively. Binding of the antibiotics piperacillin and azlocillin (**Figure 6.3A**) was confirmed by a large thermal shift of ΔT_m +17.6 ºC / +16.7 ºC for *Pa*PBP3 (**Figure 6.3B**) and ΔT_m +6.8 ºC / +5.6 ºC for *Bp*PBP3 (**Figure 6.3C**). The disparity in the extent of the thermal shifts may be a result of the difference in T_m values (approximately 11 ºC) between the two homologues.

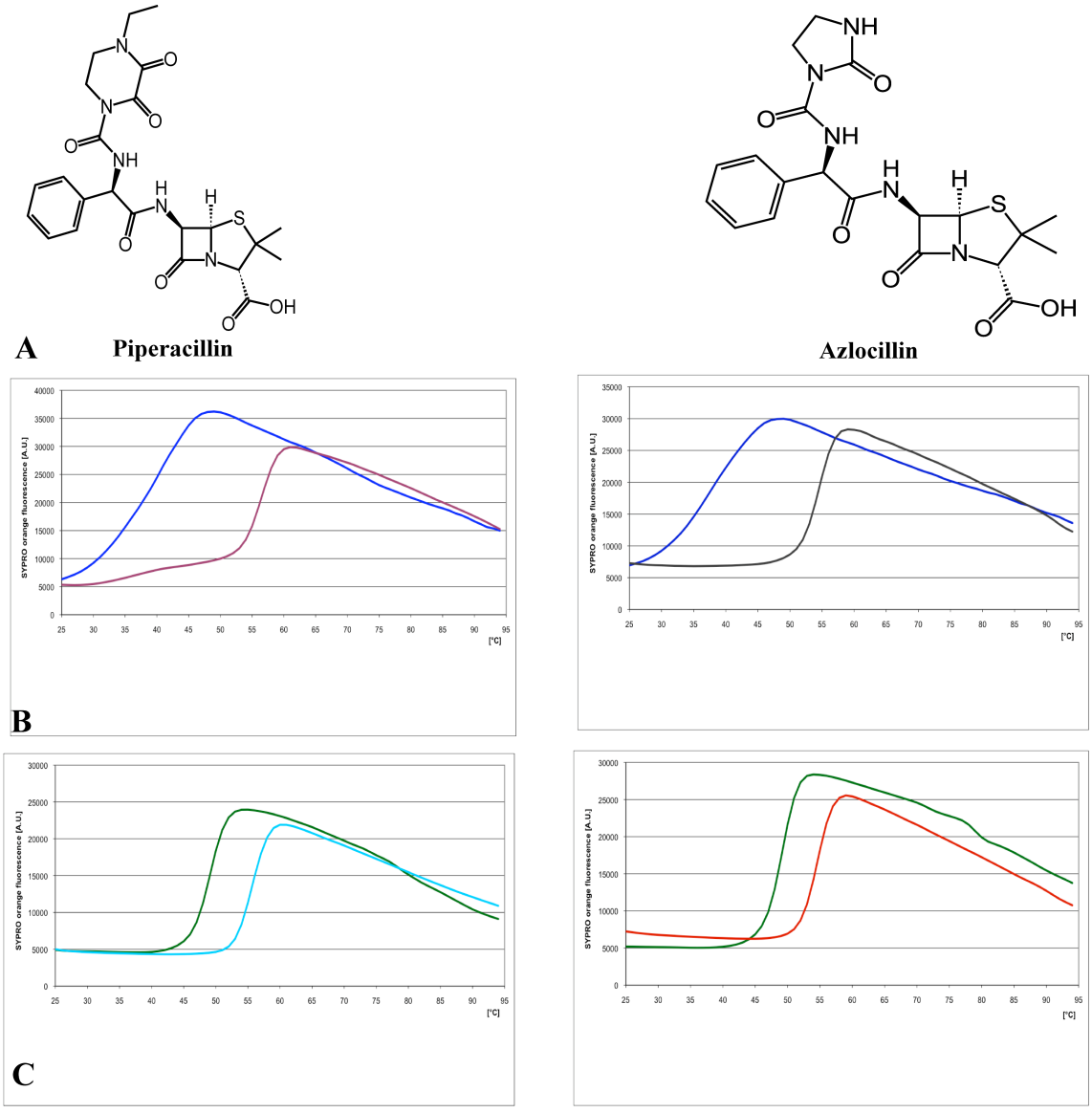


Figure 6.3 Antibiotic binds.
A) The chemical structure of piperacillin and azlocillin. **B**) *Pa*PBP3 (blue) and **C**) *Bp*PBP3 (green) are stabilised when bound to piperacillin (purple, cyan) and azlocillin (black, red). The final concentration of protein and antibiotic were 10 µM and 100 µM respectively. All experiments were carried out in triplicate and nonlinear regression was used to calculate T_m values.

Fragment screening using the Maybridge library was carried out as described in section 2.7.3. The final concentration of protein and fragment were 10 µM and 1mM

respectively. Several potential hits were identified for *Pa*PBP3, giving thermal-shifts of over 5 °C (**Figure 6.4**). A single potential hit for *Bp*PBP3 was identified giving a large thermal-shift of approximately 16 °C (**Figure 6.4**). However the melting curve of this is altered in comparison to the *Bp*PBP3 reference, with a higher initial fluorescence reading observed (**Figure 6.4**). A further experiment using BLI showed no indication of this fragment binding to *Bp*PBP3 (data not shown).

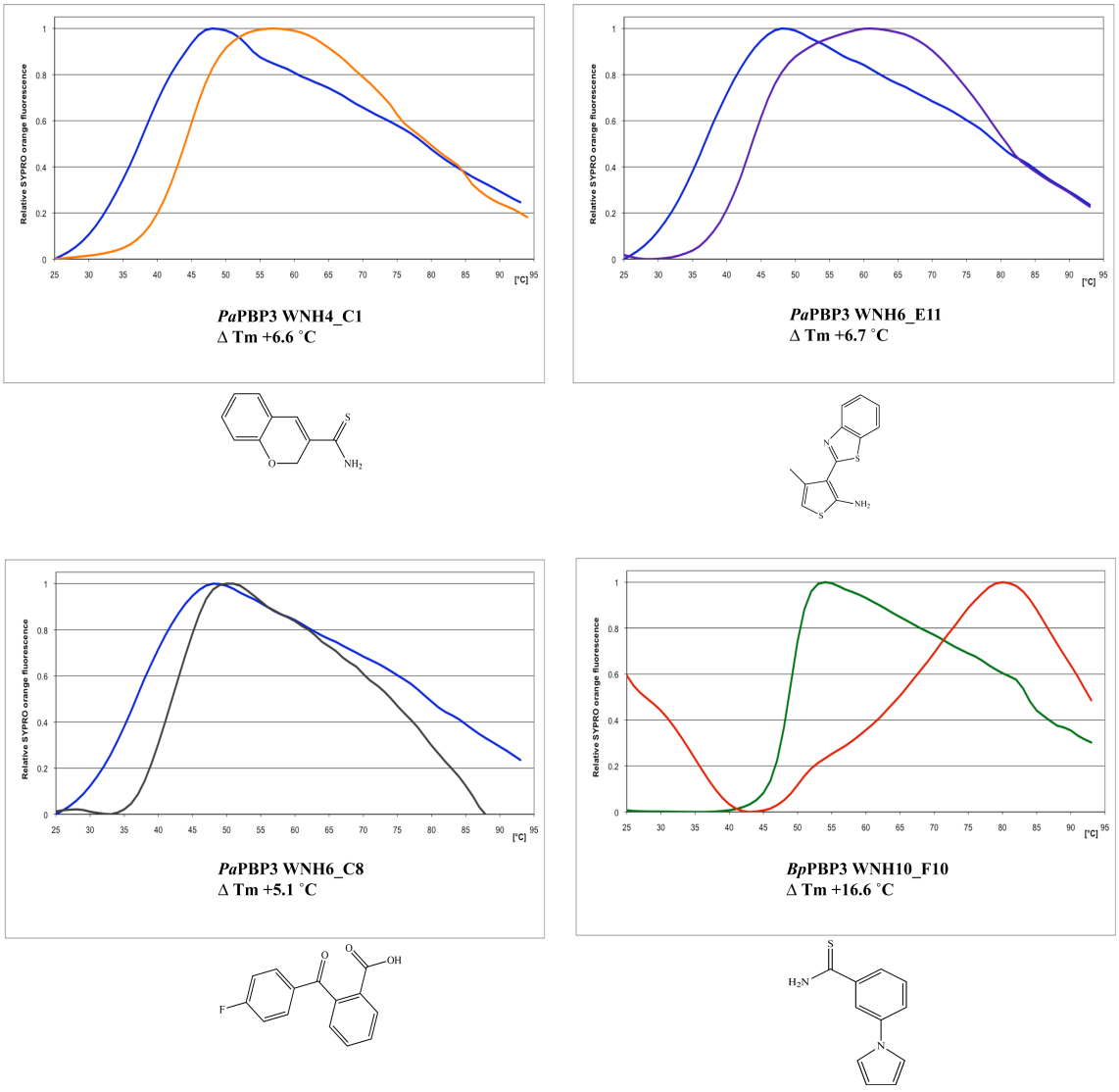
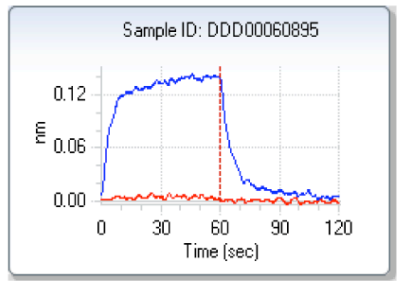


Figure 6.4. DSF fragment hits.
The thermal melting curve of *Pa*PBP3 (blue) and *Bp*PBP3 (green) are overlaid with the melting curve produced in the presence of the appropriate fragment (orange, purple, black and red). The chemical structures of the fragments are shown beneath the corresponding melting curves. The final concentration of protein and fragment were 10 µM and 1mM respectively. Screening of the Maybridge fragment library was performed once and nonlinear regression was used to calculate T_m values. Graphs are on a normalised scale to highlight the thermal-shifts observed.

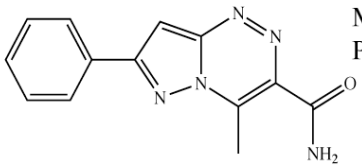
6.4.2 BLI and SPR

To increase the pool of potential fragment hits binding to *Bp*PBP3, screening using BLI (Wartchow *et al*., 2011) was carried out with biotinylated protein as described in section 2.7. BLI is a label free optical analytical technique used to analyse biomolecular interactions in real time (see section 2.7.5). Reference sensors were blocked with biocytin and used to correct for optical interference and variation in buffer composition between wells.

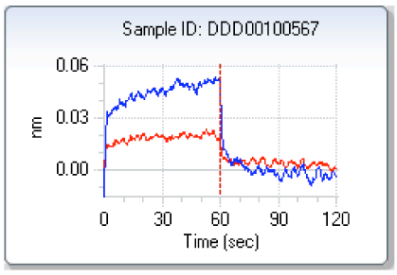
The sensograms from the BLI experiment (**Figure 6.5**) plots the change in the interference pattern of white light (representing the change of thickness at the biosensor tip; y-axis) against time (x-axis; Wartchow *et al*., 2011; see section 2.7.5). Several potential hits were identified, where a fast on/off response was observed (**Figure 6.5**). This is a typical property of fragments due their weak binding affinities (Ciulli and Abell, 2007; Scott *et al*., 2012). Two of these hits were purchased from Maybridge for co-crystallisation.



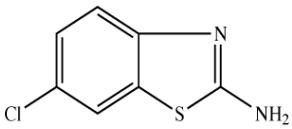
DDD00060895



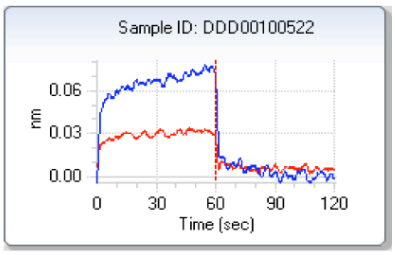
Manufacturer: Maybridge
Product Code: KM00216SC



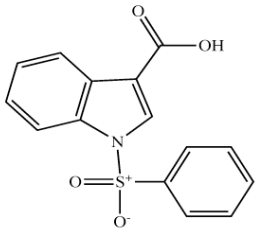
DDD00100567



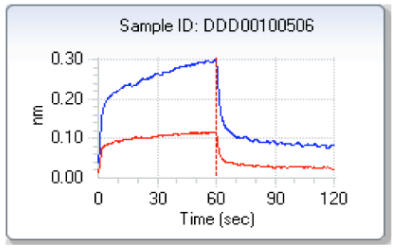
Manufacturer: Maybridge
Product Code: BTB04800SC



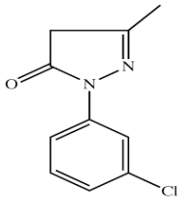
DDD00100522



Manufacturer: Maybridge
Product Code: CC03101SC



DDD00100506



Manufacturer: Maybridge
Product Code: SB00654SC

Figure 6.5 BLI sensograms.
Examples of the potential fragment hits identified, giving larger responses (blue) in comparison to the reference biosensors (red).

Screening of *Bp*PBP3 using SPR (Hoa *et al*., 2007; see section 2.7.4) was carried out to compare the results with those obtained using BLI. Dr. Leah Torrie (University of Dundee) performed this screening using a Biacore T100 (GE Healthcare) with biotinylated *Bp*PBP3 (see section 2.7) immobilised onto a streptavidin sensor (GE

Healthcare) and the concentration of fragments at 150 μM. Several potential hits were identified that were not previously observed using BLI (**Figure 6.6**). The disparity in results may be due to the difference between the two techniques; one uses a dip and read format (BLI; see section 2.7.5) while the other passes the analyte across the sensor and is more sensitive to the affects of DMSO (SPR; see section 2.7.4). Furthermore, variation in the preparation and concentration of fragments may have contributed to the difference in hits observed. The fragments DDD00100522 and DDD00100506 were identified using both BLI (**Figure 6.5**) and SPR (**Figure 6.6**). However these were discounted as they appeared as potential hits in screening campaigns with other proteins (data not shown) and are therefore likely to be non-specific binders.

The identified hits are chemically similar to one another, comprising a benzene ring at the core of the molecule. This is generally fused to either a five or six member ring (**Figure 6.5 & 6.6**) and is comparable to the fragments identified using DSF with *Pa*PBP3 (**Figure 6.4**). This may mimic the benzene ring present in piperacillin / azlocillin (**Figure 6.3A**) and suggests these hits represent good starting scaffolds to develop into novel drugs. The lack of a viable enzyme assay suitable for assessing whether or not these hits have an inhibitory effect is problematic. In general, enzyme assays for PBPs require radiolabelled benzylpenicillin and would monitor its binding to PBPs (Spratt, 1977). Competition of this binding with a fragment would not yield any results as its affinity compared to the radiolabelled substrate would be insignificant. Structural information would provide validation of fragment binding while informing on the position and interactions made with its surrounding environment.

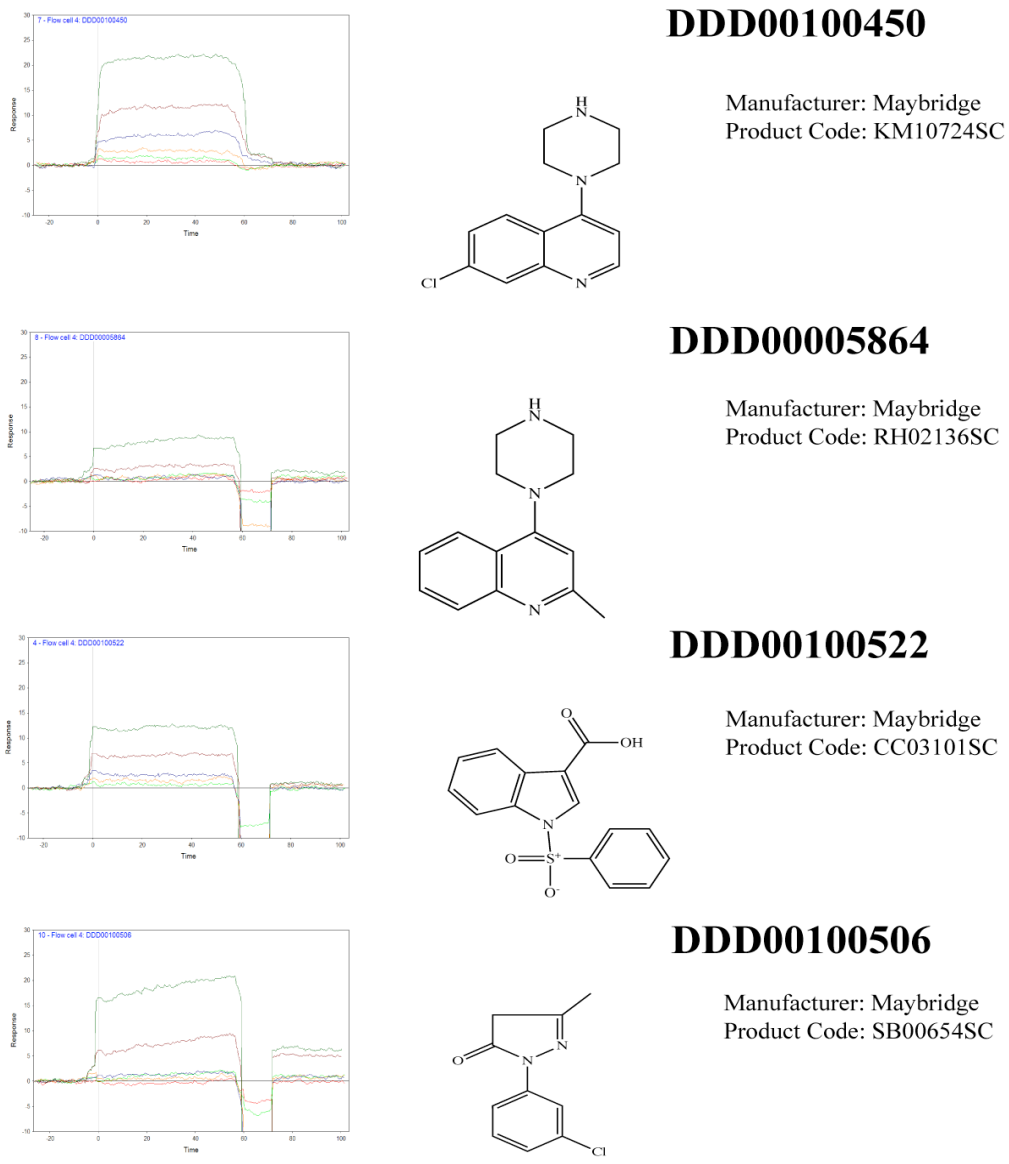


Figure 6.6 SPR sensograms.
Examples of the hits from SPR. Sensograms are coloured by concentrations.

6.4.3 Comparison of fragment screening methods

DSF has demonstrated it is a technique well-suited for detecting large changes in protein stability typically associated with substrate or inhibitor binding (**Figure 6.3B & C**). This technique was used to guide co-crystallisation of PBP3 with antibiotics; those that induced the largest increase in protein stability were selected. However, with regards to fragment screening, DSF does not seem to be a sensitive enough method for identifying the weak binding typically associated with fragments (**Figure 6.7**). With the exception of *Pa*PBP3, DSF has not generated any potential hits for *Bp*PBP3 or any

other protein screened so far using this technique (data not shown). The negative or lack of thermal shifts observed for the potential hits identified from biosensor-based methods (**Figure 6.7**) further demonstrates that this method is unsuitable for fragment screening. However with this in mind, DSF is still a convenient method that should still be considered for screening since no prior labelling or tagging of the target is required.

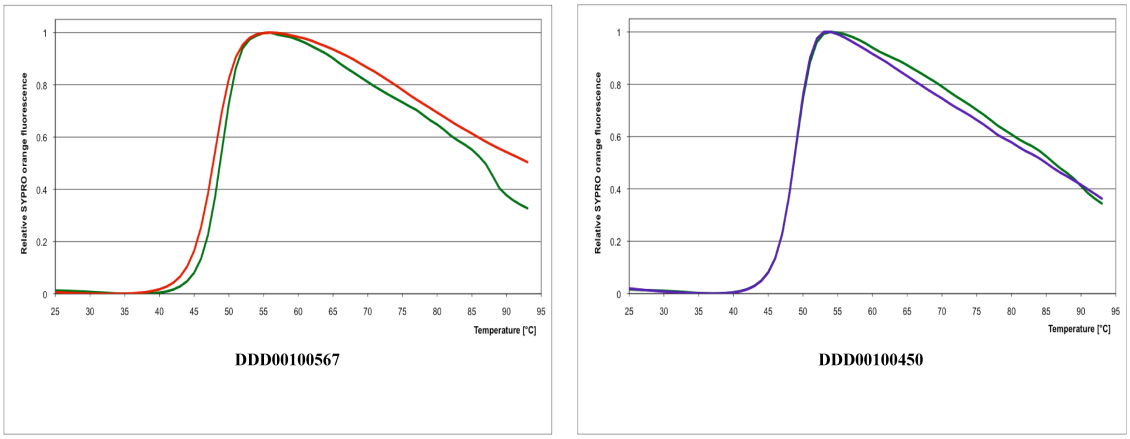


Figure 6.7 DSF not suited to fragment screening.
The thermal profile of *Bp*PBP3 (green) compared to the fragments DDD00100567 (BLI; red) and DDD00100450 (SPR; purple).

BLI screening generated several hits for *Bp*PBP3, indicating this is a technique more suited to identifying fragments. BLI has the advantage over SPR screening in that the technique is less sensitive to the effects of DMSO and provides results in a timely manner that is comparable to DSF, but without compromising sensitivity. Furthermore BLI does not require microfluidics, resulting in little maintenance and a lower cost of the system. However, SPR screening provides additional data, such as kinetic constants (K_d, K_{on}, K_{off}), for assessing fragment hits (Neumann *et al.*, 2007). Screening of fragments at several concentrations would need to be carried out to obtain such information using BLI (Wartchow *et al.*, 2011).

The Maybridge Ro3 (Maybridge) and the DDU (University of Dundee) fragment libraries were used for DSF and biosensor-based screening respectively. Both

libraries are comprised of small molecules adhering to a 'rule of three' framework (Congreve *et al*., 2003; see section 5.3) and are designed to maximise diversity while minimising over-representation of scaffolds. Analyses of results from BLI screening campaigns reveal several molecules appearing as hits for different drug targets (data not shown). This suggests a number of molecules within the DDU library are non-specific. A tailored library is therefore needed for fragment screening to remove undesirable molecules, such as those that exhibit non-specific binding. The DDU library comes close to this, however this library needs to be further improved using data generated from screening campaigns to guide the exclusion/inclusion of molecules. This would result in an evolving library optimised for screening using biosensor-based methods.

6.5 Crystallisation screening

The crystallisation of apo *Pa*PBP3 and in complex with antibiotics (carbenicillin and ceftazidime) were successfully carried out by the OPPF (Sainsbury *et al*., 2011). Attempts by the OPPF to co-crystallise *Pa*PBP3 with the fragments WNH4_C1, WNH6_E11 and WNH6_C8 identified using DSF (**Figure 6.4**) proved unsuccessful, prohibiting validation as true fragment hits.

Crystallisation screening of *Bp*PBP3 was carried out as described in section 2.4 in the absence and presence of antibiotics (piperacillin and azlocillin). Crystal trays were set up at concentrations between 10 and 15 mg mL^{-1} of protein and between 100 μM and 1 mM antibiotic at 20 °C. In addition, co-crystallisation with the fragments DDD00060895 and DDD00100567 identified using BLI (**Figure 6.5**) were carried out with 1 mM fragment. However, no crystals of *Bp*PBP3 were obtained to validate these as true hits.

6.6 Discussion

The biosynthesis of peptidoglycan (**Figure 6.1**) is an essential process for bacteria, providing structural strength and integrity to the cell (Sauvage *et al*., 2008). The enzymes involved in the polymerisation (glycosyltransferase) and cross-linking (transpeptidase) of peptidoglycan are the targets of β-lactam antibiotics and are termed penicillin-binding proteins (PBPs; Sauvage *et al*., 2008; Spratt and Cromie, 1988). PBPs can be classed into a high molecular mass (HMM) and low molecular mass (LMM) group (Sauvage *et al*., 2008), with HMM PBPs essential for cell survival (Spratt and Cromie, 1988) and LMM PBPs having little or no effect (Korsak *et al*., 2010). The HMM PBPs can be further divided into class A and B enzymes, with the class A members being bi-functional (glycosyltransferase and transpeptidase activity) and class B members being mono-functional (transpeptidase activity; Sauvage *et al*., 2008).

PBP3 is a HMM class B enzyme that is essential for cell division (Botta and Park, 1981; Spratt, 1975; Typas *et al*., 2011), with the deletion of the gene encoding PBP3 resulting in filamentation (Botta and Park, 1981; Denome *et al*., 1999). This enzyme is a validated target of β-lactam antibiotics such as piperacillin and azlocillin (Spratt and Cromie, 1988). The present study of PBP3 from *P. aeruginosa* and *B. pseudomallei* provides information on the production of soluble protein and protein-ligand interactions. These homologues share 40 % sequence identity, with many of the essential residues involved in catalysis conserved (**Figure 6.8A**). The crystal structures of apo *Pa*PBP3 (PDB 3OC2 & 3PBN) and in complex with antibiotics (**Figure 6.8B**; PDB 3OCN, 3OCL & 3PBO) represent the first structures of PBP3 to be determined (Han *et al*., 2010; Sainsbury *et al*., 2011).

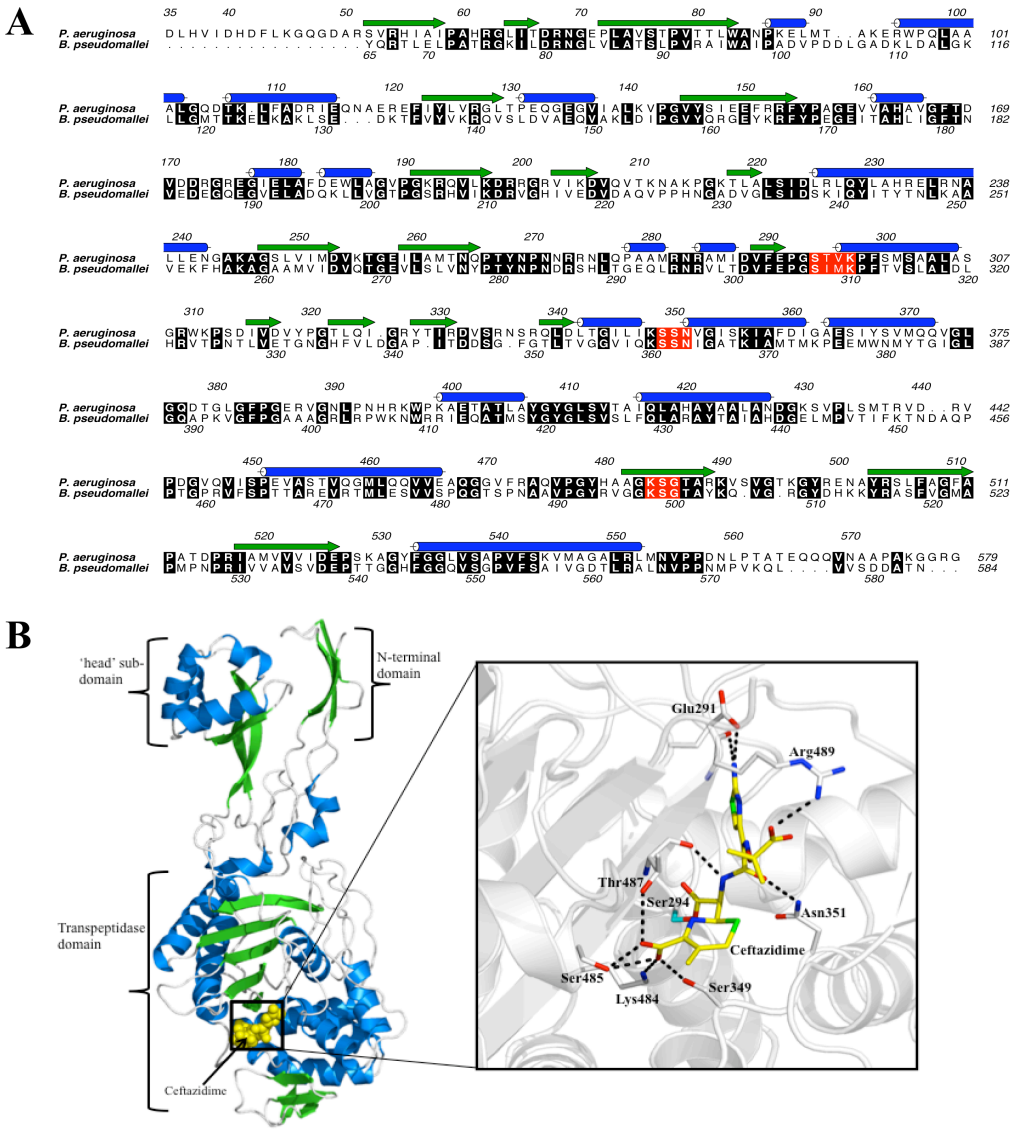


Figure 6.8 *Pa*PBP3 alignment and structure.
A) Sequence alignment of *Pa*PBP3 (NCBI reference sequence: NP_253108, PA4418) with *Bp*PBP3 (Uniprot Q63QJ1). The three structural motifs conserved among PBPs are highlighted in red, with α-helices and β-sheets coloured blue and green respectively. **B**) The structure of *Pa*PBP3 with the antibiotic ceftazidime (yellow) bound in the active site (PDB 3OCN; Sainsbury *et al*., 2011).

The information these structures provide will aid the development of novel drugs by exploiting structural features. For example, the appearance of a small binding pocket (**Figure 6.9**) resulting from conformational changes induced by inhibitor binding gives rise to the possibility of targeting this pocket to restrict the conformational flexibility of the PBP3 active site (Sainsbury *et al*., 2011). If this small pocket is formed

upon substrate binding, then a drug targeting this position has the potential to lock the active site into a single conformation, preventing further substrate binding.

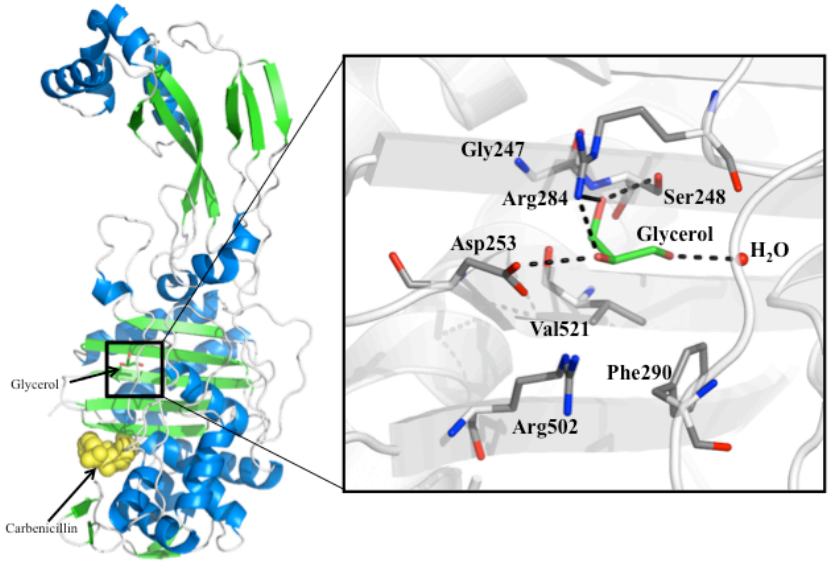


Figure 6.9 *Pa*PBP3 small binding pocket.
Conformational changes resulting from carbenicillin binding in the active site of *Pa*PBP3 (PDB 3OCL) creates a small binding pocket occupied by a well-ordered glycerol molecule (Sainsbury *et al*., 2011).

Biosensor-based screening identified a number of hits for *Bp*PBP3 (**Figure 6.5 & 6.6**). The chemical fragments all contain a benzene ring at the core of the molecule and possibly mimic the same chemical group in, for example, piperacillin. This indicates these fragments may represent starting scaffolds for novel drugs to be developed. These hits could not be validated as co-crystals could not be produced, nor could their potential inhibitory affects be ascertained due to a lack of a suitable enzyme assay. However, the fragments may be characterised in future using isothermal titration calorimetry (ITC) or BLI to determine their thermodynamic and kinetic parameters (Drinkwater *et al*., 2010; Wartchow *et al*., 2011). This information would provide confirmation of binding in the absence of an enzyme assay and structural data.

CHAPTER 7

RESULTS AND DISCUSSION: *Pseudomonas aeruginosa* INOSINE 5'-MONOPHOSPHATE DEHYDROGENASE

7.1 Background

First characterised in 1957 (Magasanik *et al*., 1957), inosine 5'-monophosphate dehydrogenase (IMPDH; EC 1.1.1.205) is an enzyme present in nearly all living organisms and catalyses the first committed step in the biosynthesis of guanine nucleotides (Hedstrom, 2009). This enzyme converts inosine 5'-monophosphate (IMP) to xanthosine 5'-monophosphate (XMP), using the reduction of NAD^+ to NADH to drive the reaction forward (**Figure 7.1A**; Hedstrom *et al*., 2011). The reaction occurs first with a catalytic cysteine residue attacking the C2 position of the IMP substrate and undergoing a dehydrogenase reaction that expels a hydride to the NAD^+ cofactor to form NADH and the covalent product intermediate E-XMP* (**Figure 7.1A**; Hedstrom *et al*., 2011). The formation of NADH leads to its expulsion from the dinucleotide binding site and the movement of a mobile flap that occupies the now empty binding site. An arginine residue then acts as a general base catalyst, allowing the hydrolysis reaction to take place to convert the E-XMP* intermediate into XMP (**Figure 7.1B;** Hedstrom *et al*., 2011).

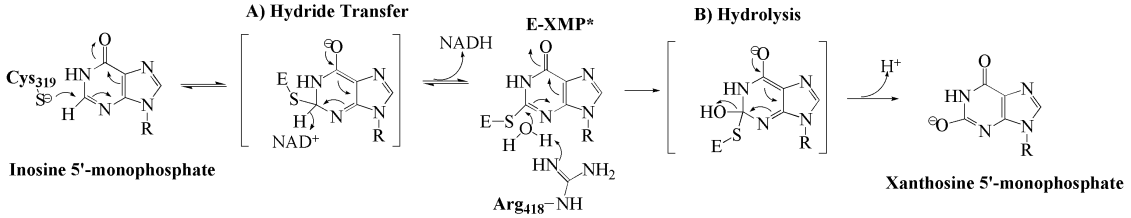


Figure 7.1 Proposed IMPDH mechanism.
The IMPDH mechanism consists of two distinct reactions; **A**) a dehydrogenation and **B**) a hydrolysis reaction. Residue numbering is based on *Tritrichomonas foetus* IMPDH (Whitby *et al*., 1997). Figure adapted from (Hedstrom *et al*., 2011).

The XMP product can be converted to the nucleotide guanine 5'-monophosphate (GMP), which is essential for proliferating cells (Jackson *et al*., 1975). Through the action of multiple enzymes, GMP provides the critical precursors for glycosylation,

DNA (dGTP) and RNA (GTP) synthesis (**Figure 7.2**; Allison and Eugui, 2000; Hedstrom, 2009; Sintchak and Nimmesgern, 2000). Other important roles of guanine nucleotides include signal transduction pathways and providing the co-factor for G-proteins (Allison and Eugui, 2000). Rapidly proliferating cells, such as cancer cells or cells involved in microbial infection, have a high demand for guanine nucleotides (Hedstrom, 2009; Hedstrom *et al*., 2011). This demand in most cases cannot be met by the occurrence of salvage pathways that may be present in the organism (**Figure 7.2**; Allison and Eugui, 2000; Hedstrom *et al*., 2011). Genetic evidence exists highlighting the importance of IMPDH with knock out studies carried out in several bacterial species indicating the gene encoding IMPDH is essential (de Berardinis *et al*., 2008; Kobayashi *et al*., 2003; Liberati *et al*., 2006). IMPDH can therefore be considered a potential drug target as inhibition can have an effect on proliferating cells (Hedstrom *et al*., 2011).

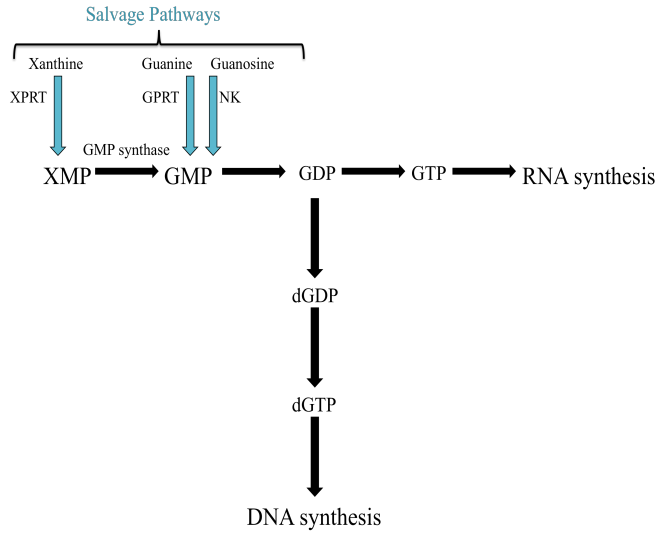


Figure 7.2 The importance of XMP. XMP is converted to GMP to make precursors for DNA/RNA synthesis. GPRT; guanine phosphoribosyl transferase, NK; Nulceoside kinase, XPRT; xanthine phosphoribosyl transferase Figure adapted from (Sintchak and Nimmesgern, 2000).

7.2 Aims

The aims of this project were to assess *P. aeruginosa* IMPDH (*Pa*IMPDH) by producing purified protein and developing both a supply of crystals for structural studies and an enzyme assay to characterise *Pa*IMPDH. In addition, identifying chemical fragments (Ciulli and Abell, 2007; Leach *et al*., 2006; Scott *et al*., 2012) that can be used as novel starting scaffolds for further drug development were sought.

7.3 Recombinant *Pa*IMPDH expression and purification

Full length *Pa*IMPDH (residues 1 – 489) was cloned from genomic DNA into both the pET15b-TEV vector and a modified version of this vector that incorporates an N-terminal BAP sequence (see section 2.72). This was carried out by Sharon Shepherd (University of Dundee).

Recombinant *Pa*IMPDH was produced in *E. coli* BL21 (DE3) pLysS cells (Stratagene). Cultures were grown in AI media (Studier, 2005) supplemented with 50 μg L^{-1} carbenicillin as described in section 2.2.2. The protein was purified as described in section 2.3 with the His-tag still present. *Pa*IMPDH eluted from a size exclusion column as two species of approximate mass of 50 kDa and 200 kDa, corresponding to a monomer and tetramer respectively. Samples were taken for SDS-PAGE analysis (**Figure 7.3**). IMPDHs are known to form catalytically active tetramers in solution (Hedstrom, 2009), therefore fractions containing the tetrameric protein were pooled and concentrated to 22 mg mL^{-1}. A final yield of 25 mg L^{-1} of bacterial culture was obtained. The purified protein was stored at 4 °C for structural studies. Aliquots were flash frozen with the addition of 2 % glycerol for fragment screening and kinetic analysis.

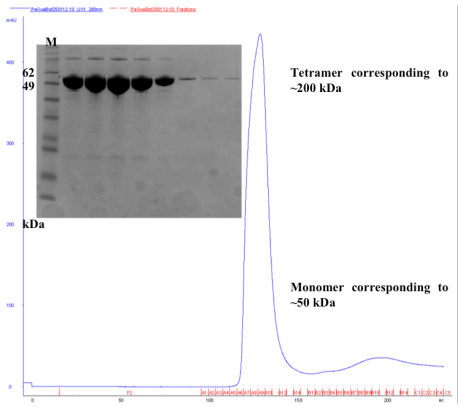


Figure 7.3 Size exclusion chromatography. Two absorbance peaks at 280 nm are observed corresponding to proteins of 50 kDa and 200 kDa. 4 mL fractions were collected.

7.4 Fragment screening

7.4.1 DSF

The thermal profile of *Pa*IMPDH (**Figure 7.4**) was determined to establish if this enzyme is amenable to DSF (Niesen *et al*., 2007; Pantoliano *et al*., 2001; see section 2.7.3) prior to commencing fragment screening. The thermal profile of *Pa*IMPDH is composed of two distinct curves with a T_m of 46.5 °C and 78.7 °C. To determine if each melting curve correlates with either the substrate binding site or the dinucleotide binding site, a number of DSF assays were carried out in the presence of ligands. Analysis of the results indicates a shift of ΔT_m +2 °C in the second melting curve as a result of the substrate (IMP) and cofactor (NAD^+; **Figure 7.4A**). When IMP and NAD^+ are assayed individually, no thermal shift is observed in this melting curve (**Figures 7.4B & C**). Taken together, this suggests the ΔT_m +2 °C thermal shift observed with IMP and NAD^+ (**Figure 7.4A**) is a result of the formation of the covalent intermediate E-XMP* or the product XMP. The second melting curve with a T_m of 78.7 °C can therefore be assigned to the substrate binding site.

DSF assays carried out with NADH result in a large thermal shift of ΔT_m +18 °C in the first melting curve (**Figure 7.4D**). As noted previously, when NAD^+ is tested no thermal shift is observed (**Figure 7.4C**), thus this appears to be a result of the difference between the reduced and oxidised forms of the cofactor. The reason for this large shift is unclear. However, the reaction mechanism of IMPDH involves the movement of a mobile flap (approximately 20 residues) into the dinucleotide binding site (**Figure 7.1B**; Hedstrom *et al*., 2011). It is possible the large thermal shift observed is linked to this, with the mobile flap inducing thermal stability in the enzyme. It is therefore possible to assign the first melting curve, with a T_m of 46.5 °C, to the dinucleotide binding site.

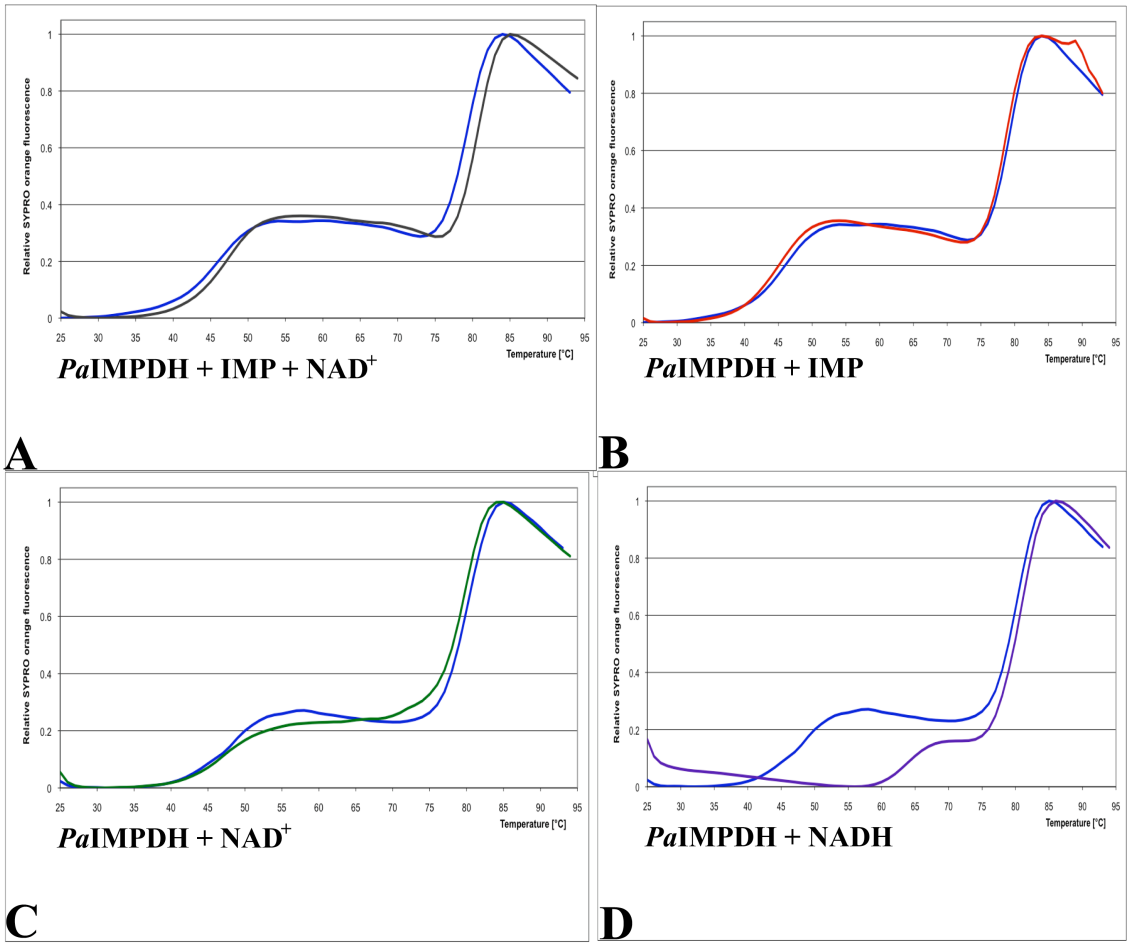


Figure 7.4 Thermal profile of *Pa*IMPDH with ligands.
The stability of *Pa*IMPDH in the presence of ligands was analysed with DSF. **A**) A thermal shift of ΔT_m +2 °C in the second melting curve is observed with the addition of IMP and NAD^+. **B** & **C**) This is not observed when either IMP or NAD^+ are analysed individually. **D**) A thermal shift of ΔT_m +18 °C is observed when NADH is assayed. Experiments were carried out in triplicate and nonlinear regression was used to calculate T_m values. Graphs are on a normalised scale to highlight the thermal-shifts observed.

With the thermal profile of *Pa*IMPDH established, fragment screening using DSF was carried out with the Maybridge fragment library as described in section 2.7.3. The screening was carried out in the presence of NAD^+ to occupy the dinucleotide binding site and thus prevent fragments binding to this site. The premise for this was to promote the finding of fragments that bind to the substrate site as the dinucleotide site is a common requirement of many enzymes. Analysis of the results showed no positive thermal shifts to indicate fragment binding. It is likely that the low binding affinity of fragments (see section 5.3) is below the threshold that can be detected using DSF.

7.4.2 BLI

Fragment screening using BLI (Wartchow *et al.*, 2011) was carried out with biotinylated *Pa*IMPDH as described in section 2.7. As with DSF fragment screening, NAD^+ was present to occupy the dinucleotide site and promote binding to the substrate site. Several potential hits were identified, where a fast on/off response was observed (**Figure 7.5**). These fragments were purchased from Maybridge for further analysis.

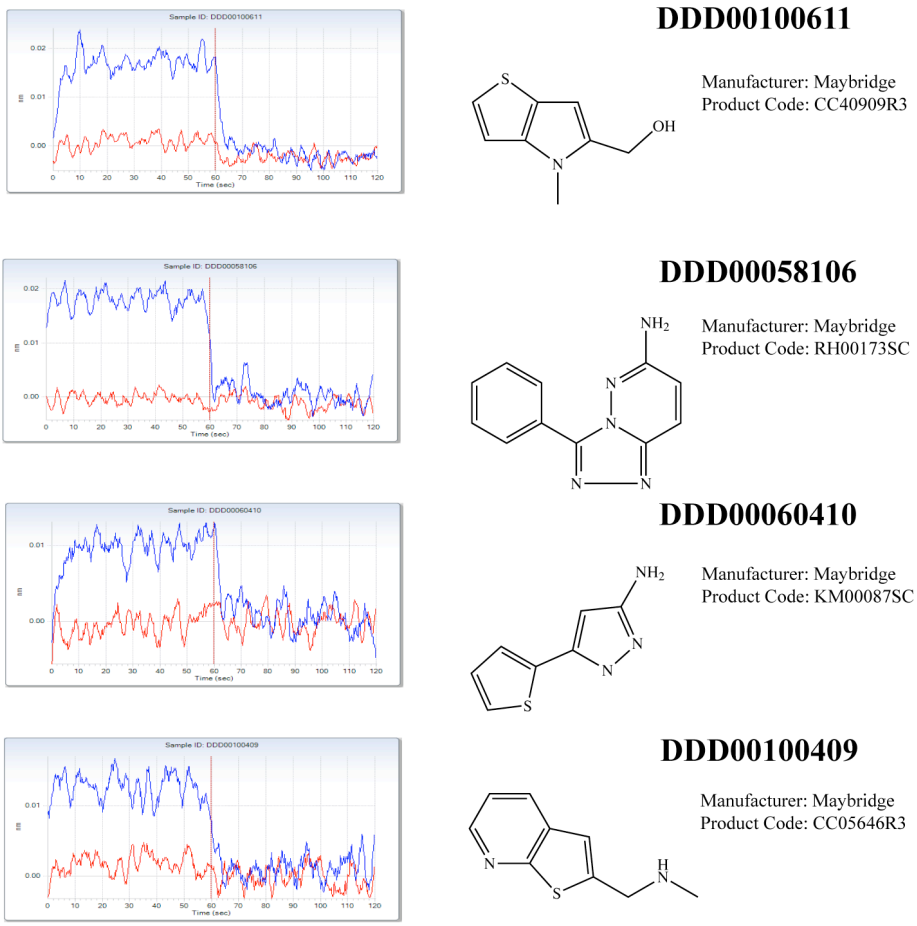


Figure 7.5 BLI sensograms.
Fragment screening using BLI resulted in a number of potential binding fragments being identified for further analysis. Reference sensors are coloured red.

7.5 Enzyme kinetics

To assess the potential fragment hits identified using BLI, enzyme assays were carried out as described in section 2.8. *Pa*IMPDH has a specific activity of 1.0 µmol min^{-1} mg^{-1} and calculated K_m values for NAD^+ and IMP of 197 (± 10) µM and 3.3 (± 0.2) mM respectively (**Figure 7.6A & B**). The K_m value for the substrate is unexpectedly distinct

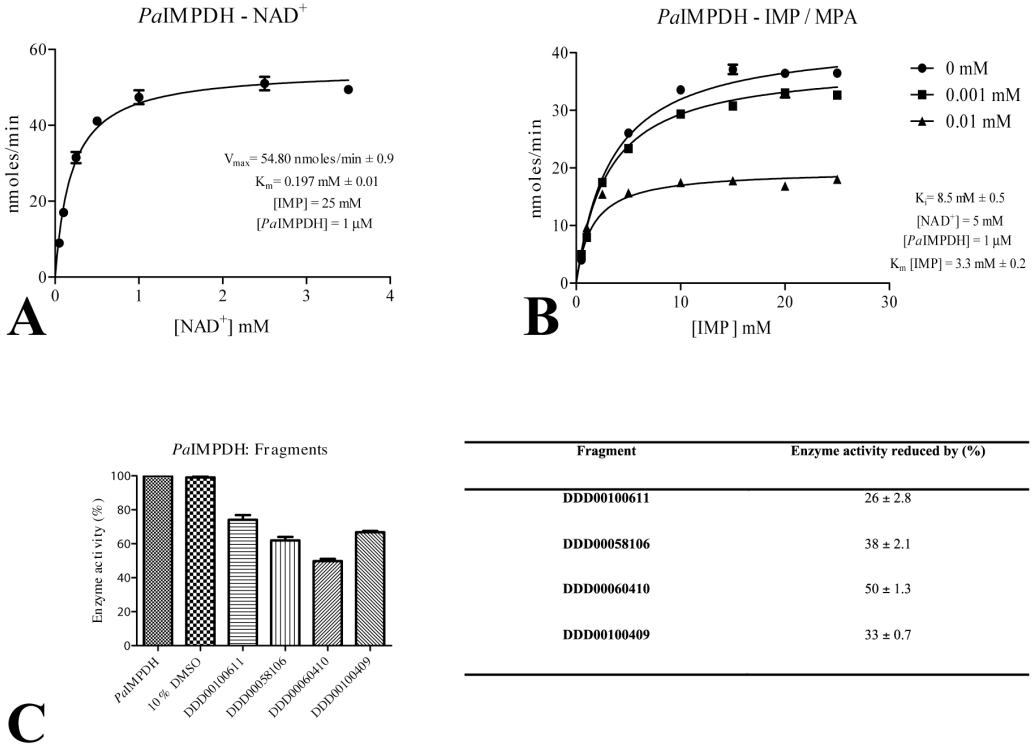
from other IMPDH homologues (**Table 7.1**; Carr *et al*., 1993; Hager *et al*., 1995; Hedstrom *et al*., 2011). It is unclear as to why such a low substrate affinity is observed for *Pa*IMPDH (**Table 7.1**), however this result was consistent across multiple experiments carried out on different occasions with different batches of protein (data not shown). The K_m for the cofactor is higher than the reported values of other bacterial homologues (with the exception of *Helicobacter pylori*) and is comparable to the values reported for the homologues in parasites (**Table 7.1**). In general, bacterial IMPDH have lower affinity for the substrate and cofactor compared to human homologues (Hedstrom, 2009; Zhang *et al*., 1999a). This difference in kinetic properties can be exploited to achieve selective inhibition (Hedstrom, 2009; Zhang *et al*., 1999a). The inhibitor mycophenolic acid (MPA; Allison and Eugui, 2000; Bentley, 2000; Carter *et al*., 1969) is an example of a selective inhibitor as it displays selectivity towards eukaryotic IMPDHs. This inhibitor is 1000 times more potent against human IMPDH compared to prokaryotic IMPDHs (**Figure 7.6B**, **Table 7.1**; Allison and Eugui, 2000; Hedstrom, 2009; see section 2.8).

Table 7.1 IMPDH kinetic parameters.

IMPDH	K_m IMP (μM)	K_m NAD^+ (μM)	K_i MPA (nM)
Human Type I	14-18	42-70	11-33
Human Type II	4-9	6-32	7-14
Aerobacter aerogenes	60	800	No data
Borelia burgdorferi	30	1100	8000
Cryptosporidium parvum	29	150	9300
E. coli	61	2000	~20,000
Helicobacter pylori	18	73	No data
Mycobacterium tuberculosis	78	1000	62,000
P. aeruginosa	**3300**	**197**	**8500**
Streptococcus pyogenes	62	1180	> 10,000
T. foetus	1.7	150	9,000

Edited from (Hedstrom *et al*., 2011).

Single point percentage inhibition with the fragments DDD00100611, DDD00058106, DDD00060410 and DDD00100409 were carried out at 1mM (dissolved in 100 % DMSO) as described in section 2.8. Each fragment reduced enzyme activity by approximately 25 to 50 % (**Figure 7.6C**). However it should be noted that upon addition of the fragment DDD00060410 to the reaction mixture, precipitation became visible and may have affected the results.



Fragment	Enzyme activity reduced by (%)
DDD00100611	26 ± 2.8
DDD00058106	38 ± 2.1
DDD00060410	50 ± 1.3
DDD00100409	33 ± 0.7

Figure 7.6 *Pa*IMPDH kinetics.
The K_m for the cofactor (**A**) and substrate (**B**) were calculated. The known IMPDH inhibitor, MPA, was tested to measure its effect on *Pa*IMPDH (**B**). The fragments identified from BLI were tested at a single concentration of 1 mM to see if an inhibitory effect could be observed (**C**). All experiments were carried out in triplicate and analyses of data were carried out using Prism (GraphPad Software).

7.6 Crystallisation and data collection

The crystal structures of *Pa*IMPDH in complex with the fragments identified in section 6.4.2 were required to assess them further. Thus conditions that promote co-crystallisation of *Pa*IMPDH with ligands (**Figure 7.7, Table 7.2**) were sought by

screening multiple conditions (see section 2.4). This was carried out at 20 °C using the sitting drop vapour diffusion method (see section 2.4) with 9 mg mL^{-1} of protein solution. Ligand soaking experiments with apo crystals (**Table 7.2**) were not carried out.

Table 7.2 *Pa*IMPDH crystallisation.

Ligand	Crystals	Condition / Days / Status	Figure
Apo	Yes	0.2 M sodium acetate, 40 % MPD, 35 days, dataset measured to 2.25 Å	6.7A
250 μM NAD^+	No		
250 μM NADH	Yes	10 % PEG 8000, 0.1 M HEPES pH 7.5, 8 % Ethylene glycol, 38 days, diffraction to 3.6 Å	6.7B
2 mM IMP / 250 μM NAD^+	Yes	10 % PEG 6000, 0.1 M HEPES pH 7.0, 8 days, no diffraction	6.7C
1 mM mycophenolic acid	No		
1 mM DDD00100611	Yes	0.2 M sodium thiocyanate, 20 % PEG 3350, 6 days, no diffraction	6.7D
1 mM DDD00058106	No		
1 mM DDD00060410	No		
1 mM DDD00100409	No		

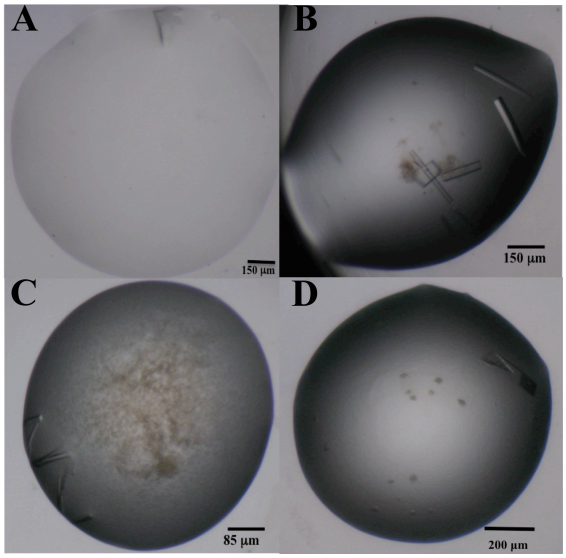


Figure 7.7 Crystals of *Pa*IMPDH.
A) Apo crystal used to measure a medium resolution dataset. **B**) Crystals with NADH, **C**) IMP/NAD^+ and the **D**) DDD00100611 fragment.

A single apo-crystal of *Pa*IMPDH was grown in 0.2 M sodium acetate, 40 % 2-methyl-2,4-pentanediol (MPD; **Figure 7.7A**) and was of diffraction quality. This was used to measure a 2.25 Å dataset in-house using a Rigaku MicroMax-007 rotating-anode X-ray generator (copper K_α, λ=1.5418 Å) coupled to an R-AXIS IV^{++} image plate detector. The high concentration of MPD present in the crystallisation condition was a suitable cryoprotectant. Data were indexed and integrated using *XDS* (Kabsch, 2010), with space group *I*4 and unit cell lengths *a*=*b*= 115.5 Å and *c*= 56.4 Å (**Table 7.3**). Data were scaled and the space group suggested using *SCALA* and *POINTLESS* (Evans, 2006) from the *CCP4* program suite (Winn *et al*., 2011). The asymmetric unit consists of a single polypeptide chain with an estimated solvent content of 60 % and a V_m (Matthews, 1968) of 2.94 $Å^3$ Da^{-1}.

7.6.1 Molecular replacement and model building

A search of the PDB using the amino acid sequence of *Pa*IMPDH showed 54 % sequence identity with *Bacillus anthracis* IMPDH (PDB 3TSB). The in-house data were used to solve the structure by molecular replacement using *PHASER* (McCoy *et al*., 2007), giving a Z-score of 23.4. The initial model consisted of 261 residues, with a correlation coefficient of 0.75 and R_{work} and R_{free} values of 37.2 % and 41.8 % respectively. Subsequent model building in *COOT* (Emsley *et al*., 2010) extended this to 293 residues, with a correlation coefficient of 0.96 and R_{work} and R_{free} values improving to 16.1 % and 19.8 % respectively. Refinement calculations were performed using *REFMAC5* (Murshudov *et al*., 2011) and translation/libration/screw analysis (TLS) was applied (Painter and Merritt, 2006). The refinement proceeded with the incorporation of water molecules and a number of side chains with dual rotamer conformations. The refinement was terminated when there were no significant changes in R_{work} and R_{free} values and inspection of the difference map suggested that no further

corrections or additions were justified. The geometry and quality of the model was validated using *MOLPROBITY* (Chen *et al*., 2010) and the subsequent Ramachandran plot reveals 99.7 % of residues were in the allowed region. A single residue was found in the disallowed region and disorder was evident at several positions where it was not possible to interpret diffuse and weak electron density. Consequently a large number of residues are absent from the final model. The polypeptide chain consists of residues 1 – 91, 205 – 370 and 426 – 467. Crystallographic statistics are given in **Table 7.3**.

Table 7.3 Crystallographic statistics.

	***Pa*IMPDH**
Space group (Å)	*I*4
a, b, c (Å)	115.5, 115.5, 56.4
Resolution[a] (Å)	19.1 – 2.25 (2.37 – 2.25)
No. reflections recorded	86989 (12454)
Unique reflections	16612 (2476)
Completeness (%)	94.6 (97.0)
Multiplicity / $\langle I/\sigma I \rangle$	5.2 (5.0) / 16.3 (4.3)
Wilson B ($Å^2$)	29.3
Residues / waters	293 / 131
R_{merge}[b] (%)	9.1 (48.9)
R_{work}[c], R_{free}[d] (%)	16.1 / 19.8
Ave. *B*-factor ($Å^2$)	
Protein / waters	22.2 / 39.3
Cruickshank DPI[e] (Å)	0.2
Ramachandran plot	
Most favoured	284
Additional allowed	8
Outliers	1
RMSD on ideal values[f]	
Bond lengths (Å) / angles (°)	0.01 / 1.2

a. Values in parentheses refer to the highest resolution shell. **b.** $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. **c.** $R_{work} = \sum_{hkl}||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. **d.** R_{free} is the same as R_{cryst} except calculated with a subset, 5 %, of data that are excluded from the refinement calculations. **e.** Diffraction Precision Index (Cruickshank, 1999). **f.** (Engh and Huber, 1991).

7.7 The *Pa*IMPDH structure

7.7.1 Overall structure

Monomers of IMPDH are known to contain two distinct domains; a catalytic domain that forms a triose-phosphate isomerase (TIM) barrel fold and a subdomain containing two cystathione β-synthase (CBS)/Bateman domains (Bateman, 1997; Hedstrom, 2009; Sintchak and Nimmesgern, 2000). The CBS domain is involved in maintaining cellular ATP/GTP concentrations (Pimkin and Markham, 2008) but is not required for enzyme activity (Nimmesgern *et al*., 1999). This domain protrudes out from the catalytic domain (**Figure 7.8A**) and is highly flexible (Hedstrom, 2009). As a result, this region is often disordered in IMPDH structures (Hedstrom, 2009).

The structure of *Pa*IMPDH displays only the TIM barrel fold of the catalytic domain (**Figure 7.8B**). A single molecule of *Pa*IMPDH is present in the asymmetric unit and the catalytic tetramer is formed by the crystallographic 4-fold axis with the operators -y,x,z, -x+1/2,-y+1/2,z-1/2 and -y+1/2,x-1/2,z-1/2 (**Figure 7.8C**). The tetramer has square planar geometry with the large missing region (114 residues) located to the corners of the tetramer (**Figure 7.8C**). *Pa*IMPDH has over 50 % sequence identity with many bacterial and parasitic homologues, including *Streptococcus pyogenes* IMPDH (PDB 1ZFJ; Zhang *et al*., 1999b). Superimposing the two structures matches 293 residues with an r.m.s.d of 1.1 Å (**Figure 7.8D**). The missing residues 92 – 204, 371 – 425 and 468 – 489 correspond to the CBS domain, mobile flap (involved in the hydrolysis reaction) and the C-terminus of *Pa*IMPDH respectively (**Figure 7.8D**).

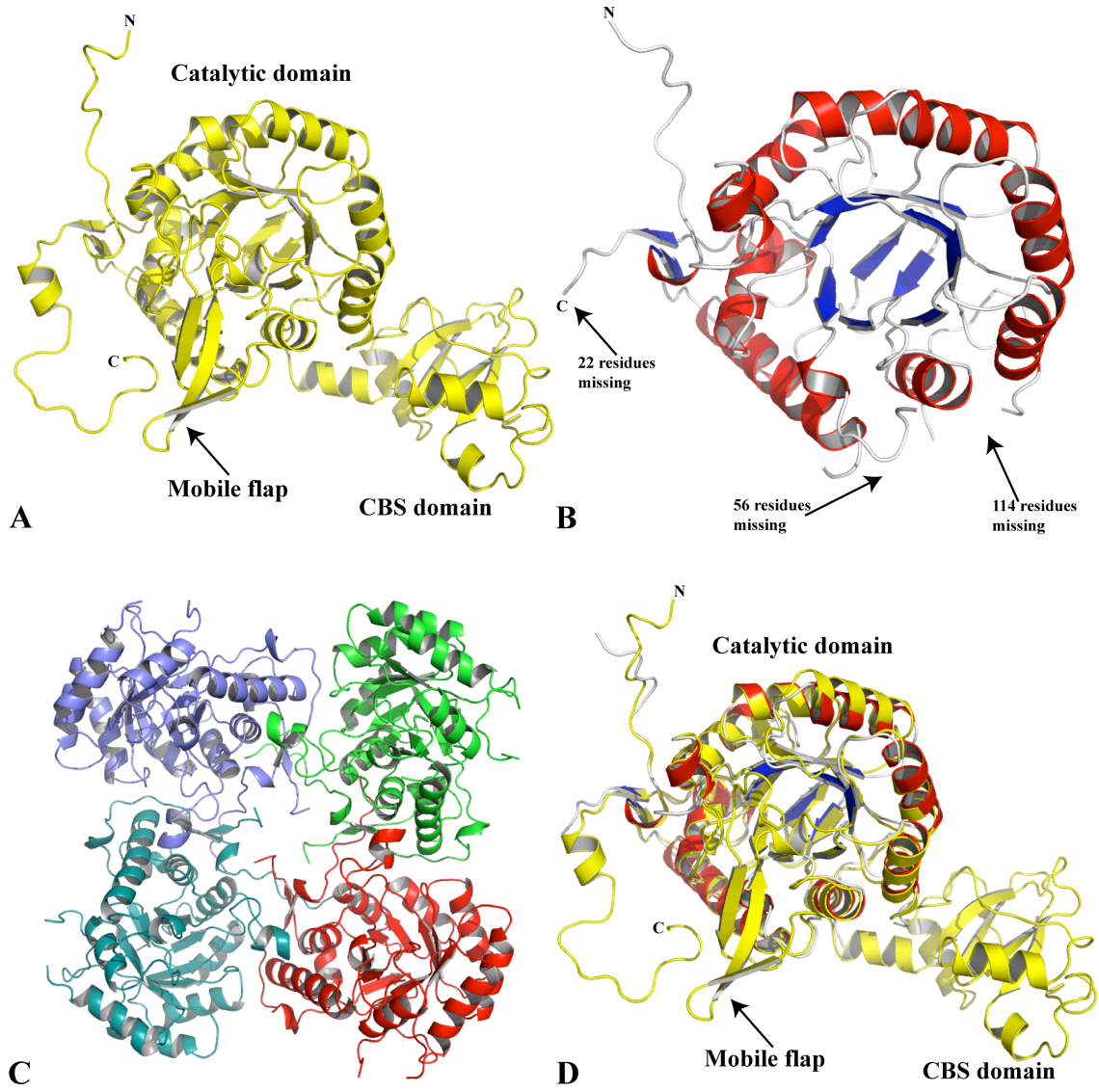


Figure 7.8 The structure of *Pa*IMPDH.
A) The structure of *S. pyogenes* IMPDH (PDB 1ZFJ; Zhang *et al*., 1999b) shows the catalytic domain, CBS domain and most of the residues comprising the mobile flap. **B**) The structure of *Pa*IMPDH shows the catalytic domain only. β-sheets and α-helices are coloured blue and red respectively. **C**) View along the crystallographic 4-fold axis showing the *Pa*IMPDH tetramer. **D**) Superimposing *Pa*IMPDH with *Sp*IMPDH highlights the missing residues in *Pa*IMPDH as corresponding to the CBS domain, mobile flap and C-terminus.

7.7.2 Active site

The active site of *Pa*IMPDH is created by a loop (residues 297 – 315) formed near the C-terminal ends of β-sheets (**Figure 7.9A**), a common feature of proteins with a TIM barrel fold (Nagano *et al*., 2002). This loop is flexible and adopts different conformations at different stages of the catalytic cycle (**Figure 7.9A**; Hedstrom, 2009; Josephine *et al*., 2010) and is often disordered in crystal structures (Hedstrom, 2009). Recent studies have shown that the active site loop controls the transition between

dehydrogenase and hydrolase activities (Josephine *et al*., 2010). This loop is fully ordered in *Pa*IMPDH with an average B-factor of 22.5 Å^2 for this region, comparable to the average for the whole protein (**Table 7.3**). It displays an "open" conformation (**Figure 7.9A**; Josephine *et al*., 2010) in contrast to the "closed" conformation observed in *Sp*IMPDH when substrate is bound (**Figure 7.9A**; Zhang *et al*., 1999b). The flexibility of the active site loop is further exemplified in the hIMPDH2 structure in complex with the inhibitor 6-Cl-IMP (PDB 1B3O & 1NFB; **Figure 7.9B**; Colby *et al*., 1999). The active site loop undergoes a conformational change that results in the short helix nearby being unwound and deformed to allow the catalytic Cys residue to attack the C6 position of the purine ring of 6-Cl-IMP (instead of the C2 position in the reaction with IMP; Colby *et al*., 1999; see section 7.1; **Figure 7.9B**).

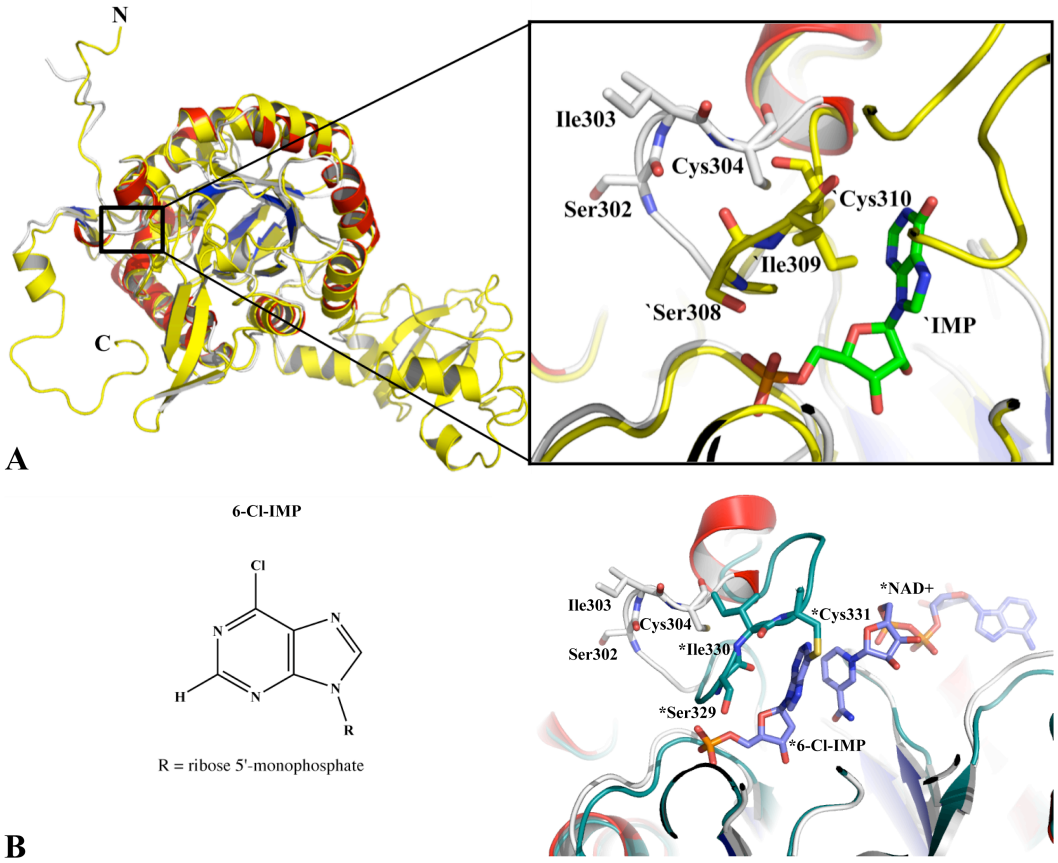


Figure 7.9 The *Pa*IMPDH active site.
A) The active site loop of *Pa*IMPDH (white) compared to *Sp*IMPDH (yellow). **B**) The chemical structure of the inhibitor 6-Cl-IMP. The active site loop of hIMPDH2 (teal; PDB 1NFB) distorts by unwinding the α-helix to allow the catalytic Cys to attack the C6 position of 6-Cl-IMP (Colby *et al*., 1999). Residues and ligands from *Sp*IMPDH and hIMPDH2 are highlighted with ` and * respectively.

7.8 Discussion

The characterisation of targets entering the drug discovery process is an important step in developing novel drugs (Bunnage, 2011; Frearson *et al*., 2007; see section 5.2). In general, IMPDHs have many of the desired attributes of an attractive drug target (see section 5.2). However, the occurrence of purine salvage pathways in organisms (**Figure 7.2**; Allison and Eugui, 2000; Hedstrom *et al*., 2011; Sintchak and Nimmesgern, 2000) complicates drug development as escape from inhibition may be possible (Hedstrom *et al*., 2011). In addition, the presence of human IMPDH homologues (Carr *et al*., 1993; Hager *et al*., 1995; Weber, 1983) makes selective inhibition of IMPDH essential for any novel drugs (Hedstrom *et al*., 2011). Two isoforms of human IMPDH exist, hIMPDH1 and hIMPDH2, sharing a sequence identity of 84 % (Carr *et al*., 1993; Hager *et al*., 1995). Yet despite such high sequence identity, selective inhibition towards hIMPDH2 has been achieved (Barnes *et al*., 2001a; Barnes *et al*., 2001b). Bacterial IMPDHs in general have 40-50 % sequence identity and have different kinetic properties compared with the human homologues (**Figure 7.10**, **Table 7.1**; Hedstrom, 2009), making selective inhibition possible (Hedstrom *et al*., 2011).

The apo structure of *Pa*IMPDH was solved and reveals the active site loop in the "open" conformation in the absence of any ligand (**Figure 7.9**). This represents only the second apo IMPDH structure (*Legionella pneumophila* IMPDH*,* PDB 3R2G) where this loop is in this conformation and is well ordered. The structure of *Borrelia burgdorferi* IMPDH (PDB 1EEP) also displays this "open" conformation, however a sulfate ion occupying the IMP phosphate binding site stabilises this conformation (McMillan *et al*., 2000). A sequence alignment of *Pa*IMPDH with IMPDH homologues shows many of the essential residues involved in catalysis are highly conserved (**Figure 7.10**, **Table 7.4**; Hedstrom, 2009). Nevertheless, variation in the residues interacting with the adenosine and pyrophosphate portions of NAD^+ (**Figure 7.10**) has been observed

(Hedstrom *et al*., 2011). This variation can be exploited to develop selective inhibitors (Hedstrom *et al*., 2011), an example of which is the inhibitors of *Cryptosporidium parvum* IMPDH (Gollapalli *et al*., 2010; Umejiego *et al*., 2008).

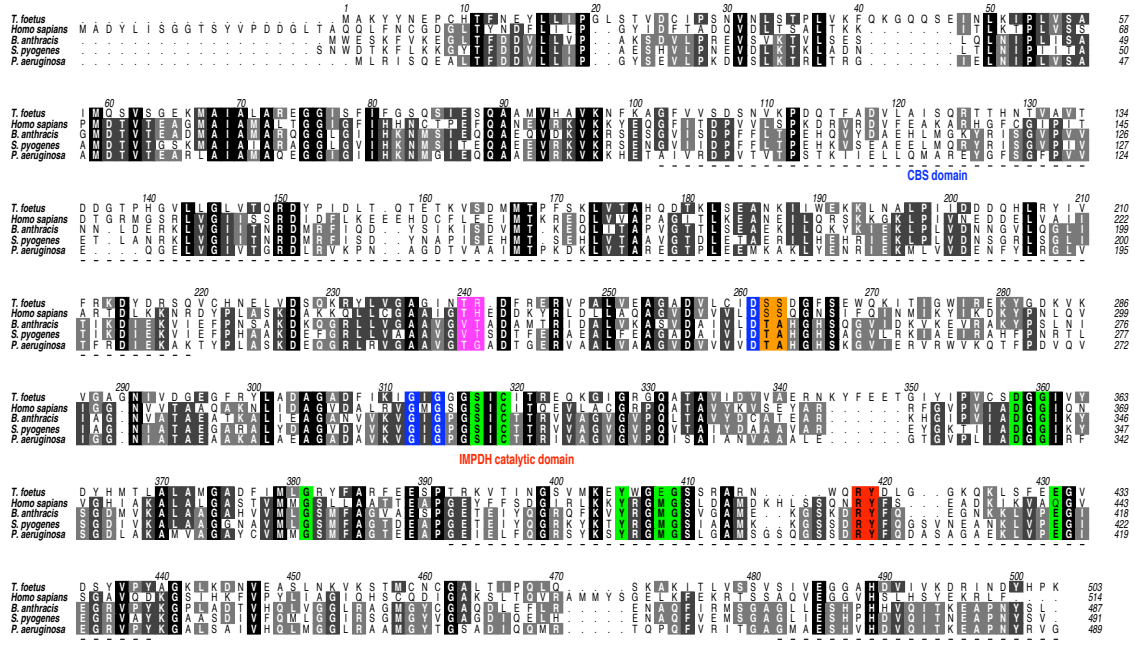


Figure 7.10 IMPDH homologues.
Sequence alignment of *Pa*IMPDH (NCBI Reference Sequence: NP_252459.1, PA3770) with homologues from *T. foetus* (GenBank ABI11203.1), *Homo sapians* (NCBI Reference Sequence: NP_000875.2), *Bacillus anthracis* (UniProt Q81W29) and *S. pyogenes* (UniProt P0C0H6). Residues involved in interactions with IMP are coloured green. Conserved residues interacting with NAD^+ are coloured blue and non-conserved residues interacting with the adenosine and pyrophosphate portions are coloured magenta and orange respectively. The two conserved residues on the mobile flap are coloured red. Residues missing in the *Pa*IMPDH structure are highlighted with a dashed line beneath them.

Table 7.4 Functional IMPDH residues.

Residue (numbering based on *T. foetus*)	Interactions / Role
Interactions with IMP:	
Cys319	Essential catalytic residue.
Ser317 / Tyr405	Forms H-bonds to the ribose phosphate via hydroxyl groups.
Gly360 / Gly381	Forms H-bonds to the ribose phosphate via main chain amides.
Asp358	Forms H-bonds to the ribose hydroxyls of IMP.
Glu408 (often a Meth) / Gly409 / Glu431 (a Gln in hIMPDH2)	Interacts with the purine ring of IMP via main chain H-bonds.
Interactions with NAD^+:	
Asp261	Forms H-bonds with ribose hydroxyls of the nicotinamide portion of NAD^+.
Gly312 / Gly314	Forms H-bonds with the carboxamide of NAD^+.
Mobile flap:	
Arg418 / Tyr419	Only residues that are completely conserved in this region. Essential for catalysis.

Fragment screening of *Pa*IMPDH was carried out in the presence of NAD^+ to identify chemical fragments that might bind in the substrate binding site (see section 7.4). Specificity could be achieved further in the drug discovery process by fragment linking/growing (Scott *et al*., 2012) into the adjacent dinucleotide binding site (Hedstrom *et al*., 2011). Several potential fragment hits were identified using BLI (**Figure 7.5**) with single point inhibition studies indicating each fragment has an inhibitory effect on the enzyme activity of *Pa*IMPDH (**Figure 7.6C**). However, no structural information about these fragments could be obtained as diffraction quality co-crystals could not be produced. The structure of *Pa*IMPDH in complex with a fragment would provide confirmation of fragment binding and provide data as to the position and interactions made by the fragment. Failure to obtain a complex structure of *Pa*IMPDH in future may be offset by mutagenesis studies. If the identified fragments truly bind to the active site of *Pa*IMPDH, then mutation of conserved residues in this region can help define the residues involved in forming the interaction. This would aid in potentially optimising the fragments into drug-like compounds. Further experiments will subsequently be required to fully characterise the inhibitory effects of these fragments on enzyme activity by calculating IC_{50} values for each. In addition, the distinct K_m value for the IMP substrate (**Figure 7.6B**; **Table 7.1**) would in future need to be verified and an explanation for the vast difference with other IMPDH homologues explained.

CHAPTER 8

REFERENCES

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC and Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst* D**66**: 213-221

Aksyuk AA, Leiman PG, Kurochkina LP, Shneider MM, Kostyuchenko VA, Mesyanzhinov VV and Rossmann MG (2009) The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO J* **28**: 821-829

Allison AC and Eugui EM (2000) Mycophenolate mofetil and its mechanisms of action. *Immunopharmacology* **47**: 85-118

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402

Arias CA and Murray BE (2009) Antibiotic-resistant bugs in the 21st century - a clinical super-challenge. *N Engl J Med* **360**: 439-443

Aschtgen M-S, Zoued A, Lloubès R, Journet L and Cascales E (2011) The C-tail anchored TssL subunit, an essential protein of the enteroaggregative *Escherichia coli* Sci-1 Type VI secretion system, is inserted by YidC. *Microbiology Open* **1**: 71-82

Aschtgen M-S, Bernard CS, De Bentzmann S, Lloubès R and Cascales E (2008) SciN is an outer membrane lipoprotein required for type VI secretion in enteroaggregative *Escherichia coli*. *J Bacteriol* **190**: 7523-7531

Aschtgen M-S, Gavioli M, Dessen A, Lloubès R and Cascales E (2010a) The SciZ protein anchors the enteroaggregative *Escherichia coli* Type VI secretion system to the cell wall. *Mol Microbiol* **75**: 886-899

Aschtgen M-S, Thomas MS and Cascales E (2010b) Anchoring the type VI secretion system to the peptidoglycan: TssL, TagL, TagP... what else? *Virulence* **1**: 535-540

Baker NA, Sept D, Joseph S, Holst MJ and McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**: 10037-10041

Ball AP, McGhie D and Geddes AM (1977) *Serratia marcescens* in a general hospital. *Q J Med* **46**: 63-71

Ballister ER, Lai AH, Zuckermann RN, Cheng Y and Mougous JD (2008) *In vitro* self-assembly of tailorable nanotubes from a simple protein building block. *Proc Natl Acad Sci USA* **105**: 3733-3738

Barnes BJ, Izydore RA, Eakin AE and Hall IH (2001a) Induction of Tmolt4 leukemia cell death by 3,3-disubstituted-6,6-pentamethylene-1,5-diazabicyclo[3.1.0]hexane-2-4-diones: specificity for type II inosine 5'-monophasphate dehydrogenase. *J Pharm Exp Ther* **298**: 790-6

Barnes BJ, Izydore RA, Eakin AE and Hall IH (2001b) Mechanism of action of the antitumor agents 6-benzoyl-3,3-disubstituted-1,5-diazabicyclo[3.1.0]hexane-2,4-diones: potent inhibitors of human type II inosine 5'-monophosphate dehydrogenase. *Int J Cancer* **94**: 275-81

Basler M, Pilhofer M, Henderson GP, Jensen GJ and Mekalanos JJ (2012) Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**: 182-186

Bateman A (1997) The structure of a domain common to archaebacteria and the homocystinuria disease protein. *Trends in Biochemical Sciences* **22**: 12-13

Beck MR, DeKoster GT, Hambly DM, Gross ML, Cistola DP and Goldman WE (2008) Structural features responsible for the biological stability of Histoplasma's virulence factor CBP. *Biochemistry* **47**: 4427-4438

Bentley R (2000) Mycophenolic Acid: a one hundred year odyssey from antibiotic to immunosuppressant. *Chem Rev* **100**: 3801-3826

Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD and Zardecki C (2002) The Protein Data Bank. *Acta Cryst* D**58**: 899-907

Bertani G (1951) Studies on lysogenesis: The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* **62**: 293-300

Bingle L, Bailey C, Pallen M and Mark J (2008) Type VI secretion: a beginner's guide. *Curr Opin Microbiol* **11**: 3-8

Bizio B (1823) Lettera di Bartolomeo Bizio al chiarissimo canonico Angelo Bellani sopra il fenomeno della polenta porporina. *Bibl Ital G Lett Sci Art (Anno VIII)* **30**: 275-295

Blake CCF, Geisow MJ, Oatley SJ, Rérat B and Rérat C (1978) Structure of prealbumin: Secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å. *J Mol Biol* **121**: 339-356

Bollag G, Hirth P, Tsai J, Zhang J, Ibrahim PN, Cho H, Spevak W, Zhang C, Zhang Y, Habets G, Burton EA, Wong B, Tsang G, West BL, Powell B, Shellooe R, Marimuthu A, Nguyen H, Zhang KYJ, Artis DR, Schlessinger J, Su F, Higgins B, Iyer R, D'Andrea K, Koehler A, Stumm M, Lin PS, Lee RJ, Grippo J, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, Chapman PB, Flaherty KT, Xu X, Nathanson KL and Nolop K (2010) Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **467**: 596-599

Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19**: 311-312

Bond CS and Schüttelkopf AW (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Cryst* D**65**: 510-512

Bönemann G, Pietrosiuk A, Diemand A, Zentgraf H and Mogk A (2009) Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *EMBO J* **28**: 315-325

Bönemann G, Pietrosiuk A and Mogk A (2010) Tubules and donuts: a type VI secretion story. *Mol Microbiol* **76**: 815-821

Bos MP, Robert V and Tommassen J (2007) Biogenesis of the Gram-negative bacterial outer membrane. *Annu Rev Microbiol* **61**: 191-214

Botta GA and Park JT (1981) Evidence for involvement of penicillin-binding protein 3 in murein synthesis during septation but not during cell elongation. *J Bacteriol* **145**: 333-340

Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, Scheld M, Spellberg B and Bartlett J (2009) Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin Infect Dis* **48**: 1-12

Boyer F, Fichant G, Berthod J, Vandenbrouck Y and Attree I (2009) Dissecting the bacterial type VI secretion system by a genome wide *in silico* analysis: what can be learned from available microbial genomic resources? *BMC genomics* **10**: 104-118

Brown D and Superti-Furga G (2003) Rediscovering the sweet spot in drug discovery. *Drug Discovery Today* **8**: 1067-1077

Bunnage ME (2011) Getting pharmaceutical R&D back on target. *Nat chem biol* **7**: 335-339

Calfee DP (2012) Crisis in hospital-acquired, healthcare-associated infections. *Annu Rev Med* **63**: 359-371

Carr SF, Papp E, Wu JC and Natsumeda Y (1993) Characterization of human type I and type II IMP dehydrogenases. *J Biol Chem* **268**: 27286-27290

Carter SB, Franklin TJ, Jones DF, Leonard BJ, Mills SD, Turner RW and Turner WB (1969) Mycophenolic acid: an anti-cancer compound with unusual properties. *Nature* **223**: 848-850

Cascales E (2008) The type VI secretion toolkit. *EMBO Rep* **9**: 735-741

Cascales E and Cambillau C (2012) Structural biology of type VI secretion systems. *Phil Trans R Soc B* **367**: 1102-1111

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS and Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystr* D**66**: 12-21

Chen Y, Wong J, Sun GW, Liu Y, Tan G-YG and Gan Y-H (2011) Regulation of type VI secretion system during *Burkholderia pseudomallei* infection. *Infect Immun* **79**: 3064-3073

Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, Salzberg AC and Huang ES (2007) Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* **25**: 71-75

Chopra I, Schofield C, Everett M, O'Neill A, Miller K, Wilcox M, Frère JM, Dawson M, Czaplewski L and Urleb U (2008) Treatment of health-care-associated infections caused by Gram-negative bacteria: a consensus statement. *Lancet infect dis* **8**: 133-139

Ciulli A and Abell C (2007) Fragment-based approaches to enzyme inhibition. *Curr Opin Biotechnol* **18**: 489-496

Colby TD, Vanderveen K, Strickler MD, Markham GD and Goldstein BM (1999) Crystal structure of human type II inosine monophosphate dehydrogenase: implications for ligand binding and drug design. *Proc Natl Acad Sci USA* **96**: 3531-3536

Congreve M, Carr R, Murray C and Jhoti H (2003) A 'rule of three' for fragment-based lead discovery? *Drug Discovery Today* **8**: 876-877

Crago AM and Koronakis V (1998) *Salmonella* InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization. *Mol Microbiol* **30**: 47-56

Cruickshank DW (1999) Remarks about protein structure precision. *Acta Cryst* D**55**: 583-601

Daefler S and Russel M (1998) The *Salmonella typhimurium* InvH protein is an outer membrane lipoprotein required for the proper localization of InvG. *Mol Microbiol* **28**: 1367-1380

Dauter Z, Dauter M and Rajashankar K (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Cryst* D**56**: 232-237

De Berardinis V, Vallenet D, Castelli V, Besnard M, Pinet A, Cruaud C, Samair S, Lechaplais C, Gyapay G, Richez C, Durot M, Kreimeyer A, Le Fèvre F, Schächter V, Pezo V, Döring V, Scarpelli C, Médigue C, Cohen GN, Marlière P, Salanoubat M and Weissenbach J (2008) A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol Syst Biol* **4**: 1-15

Denome SA, Elf PK, Henderson TA, Nelson DE and Young KD (1999) *Escherichia coli* mutants lacking all possible combinations of eight penicillin binding proteins: viability, characteristics, and implications for peptidoglycan synthesis. *J Bacteriol* **181**: 3981-3993

Desvaux M, Hébraud M, Talon R and Henderson IR (2009) Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends Microbiol* **17**: 139-145

Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang J-M, Taly J-F and Notredame C (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* **39** (Web Server issue): W13-17

Drenth J and Mesters J (2006) Principles of protein x-ray crystallography. ISBN 978-0-3873-3334-2

Drinkwater N, Vu H, Lovell KM, Criscione KR, Collins BM, Prisinzano TE, Poulsen S-A, McLeish MJ, Grunewald GL and Martin JL (2010) Fragment-based screening by X-ray crystallography, MS and isothermal titration calorimetry to identify PNMT (phenylethanolamine N-methyltransferase) inhibitors. *Biochem J* **431**: 51-61

Durand E, Zoued A, Spinelli S, Watson PJH, Aschtgen M-S, Journet L, Cambillau C and Cascales E (2012) Structural characterization and oligomerization of the TssL protein, a component shared by the bacterial Type VI and Type IVb secretion systems. *J Biol Chem* **287**: 14157-14168

Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of Coot. *Acta Cryst* D**66**: 486-501

Eneqvist T, Lundberg E, Karlsson A, Huang S, Santos CRA, Power DM and Sauer-Eriksson AE (2004) High resolution crystal structures of piscine transthyretin reveal different binding modes for triiodothyronine and thyroxine. *J Biol Chem* **279**: 26411-26416

Engh R and Huber R (1991) Accurate bond and angle parameters for x-ray protein-structure refinement. *Acta Cryst* A**47**: 392-400

English G, Trunk K, Rao VA, Srikannathasan V, Hunter WN and Coulthurst SJ (2012) New self-resistance proteins encoded within the type VI secretion system gene cluster of *Serratia marcescens*. *Mol Microbiol* Submitted.

Erlanson DA, McDowell RS and O'Brien T (2004) Fragment-based drug discovery. *J Med Chem* **47**: 3463-3482

Evans P (2006) Scaling and assessment of data quality. *Acta Cryst* D**62**: 72-82

Fauman EB, Rai BK and Huang ES (2011) Structure-based druggability assessment-identifying suitable targets for small molecule therapeutics. *Curr Opin Chem Biol* **15**: 463-468

Felisberto-Rodrigues C, Durand E, Aschtgen M-S, Blangy S, Ortiz-Lombardia M, Douzi B, Cambillau C and Cascales E (2011) Towards a structural comprehension of bacterial Type VI secretion systems: characterization of the TssJ-TssM complex of an *Escherichia coli* pathovar. *PLoS Pathog* **7**: e1002386

Felise HB, Nguyen HV, Pfuetzner RA, Barry KC, Jackson SR, Blanc MP, Bronstein PA, Kline T and Miller SI (2008) An inhibitor of Gram-negative bacterial virulence protein secretion. *Cell Host Microbe* **4**: 325-336

Fernandes P (2006) Antibacterial discovery and development - the failure of success? *Nat Biotechnol* **24**: 1497-1503

Frearson JA, Wyatt PG, Gilbert IH and Fairlamb AH (2007) Target assessment for antiparasitic drug discovery. *Trends Parasitol* **23**: 589-595

Fronzes R, Christie PJ and Waksman G (2009) The structural biology of type IV secretion systems. *Nat Rev Micro* **7**: 703-714

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD and Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook*: 571-607

Gaynes R, Edwards JR and the National Nosocomial Infections Surveillance System (2005) Overview of nosocomial infections caused by Gram-negative bacilli. *Clin Infect Dis* **41**: 848-854

Gerlach RG and Hensel M (2007) Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens. *Int J Med Microbiol* **297**: 401-415

Gollapalli DR, Macpherson IS, Liechti G, Gorla SK, Goldberg JB and Hedstrom L (2010) Structural determinants of inhibitor selectivity in prokaryotic IMP dehydrogenases. *Chemistry & Biology* **17**: 1084-1091

Hager PW, Collart FR, Huberman E and Mitchell BS (1995) Recombinant human inosine monophosphate dehydrogenase type I and type II proteins. Purification and characterization of inhibitor binding. *Biochem Pharmacol* **49**: 1323-1329

Hajduk PJ, Huth JR and Tse C (2005) Predicting protein druggability. *Drug Discovery Today* **10**: 1675-1682

Hann MM, Leach AR and Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci* **41**: 856-864

Hardie KR, Lory S and Pugsley AP (1996) Insertion of an outer membrane protein in *Escherichia coli* requires a chaperone-like protein. *EMBO J* **15**: 978-988

Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ and Jhoti H (2005) Fragment-based lead discovery using X-ray crystallography. *J Med Chem* **48**: 403-413

Hedstrom L (2009) IMP dehydrogenase: structure, mechanism and inhibition. *Chem Rev* **109**: 2903-2928

Hedstrom L, Liechti G, Goldberg JB and Gollapalli DR (2011) The antibiotic potential of prokaryotic IMP dehydrogenase inhibitors. *Curr Med Chem* **18**: 1909-1918

Hejazi A and Falkiner F (1997) *Serratia marcescens*. *J Med Microbiol* **46**: 903-912

Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC and Ala'Aldeen D (2004) Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev* **68**: 692-744

Hennebry SC (2009) Evolutionary changes to transthyretin: structure and function of a transthyretin-like ancestral protein. *FEBS J* **276**: 5367-5379

Hennebry SC, Law RHP, Richardson SJ, Buckle AM and Whisstock JC (2006) The crystal structure of the transthyretin-like protein from *Salmonella dublin*, a prokaryote 5-hydroxyisourate hydrolase. *J Mol Biol* **359**: 1389-1399

Hidron AI, Edwards JR, Patel J, Horan TC, Sievert DM, Pollock DA, Fridkin SK, NHSN and PNHSN (2008) NHSN annual update: antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006-2007. *Infect Control Hosp Epidemiol* **29**: 996-1011

Hoa X, Kirk A and Tabrizian M (2007) Towards integrated and sensitive surface plasmon resonance biosensors: a review of recent progress. *Biosensors and Bioelectronics* **23**: 151-160

Hoelder S, Clarke PA and Workman P (2012) Discovery of small molecule cancer drugs: successes, challenges and opportunities. *Molecular Oncology* **6**: 155-176

Holland IB (2010) The extraordinary diversity of bacterial protein secretion mechanisms. *Methods Mol Biol* **619**: 1-20

Holland IB, Schmitt L and Young J (2005) Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway (review). *Mol Membr Biol* **22**: 29-39

Hood RD, Singh P, Hsu F, Güvener T, Carl MA, Trinidad RRS, Silverman JM, Ohlson BB, Hicks KG, Plemel RL, Li M, Schwarz S, Wang WY, Merz AJ, Goodlett DR and

Mougous JD (2010) A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* **7**: 25-37

Hopkins AL and Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* **1**: 727-730

Hopkins A, Groom C and Alex A (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **9**: 430-431

Ireland DC, Colgrave ML and Craik DJ (2006) A novel suite of cyclotides from *Viola odorata*: sequence variation and the implications for structure, function and stability. *Biochem J* **400**: 1-12

Jackson RC, Weber G and Morris HP (1975) IMP dehydrogenase, an enzyme linked with proliferation and malignancy. *Nature* **256**: 331-333

Jancarik J and Kim SH (1991) Sparse matrix sampling: A screening method for crystallization of proteins. *J Appl Cryst* **24**: 409-411

Jani AJ and Cotter PA (2010) Type VI secretion: not just for pathogenesis anymore. *Cell Host Microbe* **8**: 2-6

Johnson TL, Abendroth J, Hol WGJ and Sandkvist M (2006) Type II secretion: from structure to function. *FEMS Microbiol Lett* **255**: 175-186

Josephine HR, Ravichandran KR and Hedstrom L (2010) The Cys319 loop modulates the transition between dehydrogenase and hydrolase conformations in IMP dehydrogenase. *Biochemistry* **49**: 10674-10681

Kabsch W (2010) XDS. *Acta Cryst* D**66**: 125-132

Kabsch W and Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**: 2577-2637

Kanamaru S, Leiman PG, Kostyuchenko VA, Chipman PR, Mesyanzhinov VV, Arisaka F and Rossmann MG (1994) Structure of the cell-puncturing device of bacteriophage T4. *Proteins* **239**: 349-363

Katz ML, Mueller LV, Polyakov M and Weinstock SF (2006) Where have all the antibiotic patents gone? *Nat Biotechnol* **24**: 1529-1533

Kelley LA and Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**: 363-371

Kerr KG and Snelling AM (2009) *Pseudomonas aeruginosa*: a formidable and ever-present adversary. *J Hosp Infect* **73**: 338-344

Kluge AF and Petter RC (2010) Acylating drugs: redesigning natural covalent inhibitors. *Curr Opin Chem Biol* **14**: 421-427

Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Débarbouille M, Dervyn E, Deuerling E, Devine K, Devine SK, Dreesen O, Errington J, Fillinger S, Foster SJ, Fujita Y, Galizzi A, Gardan R, Eschevins C, Fukushima T, Haga K, Harwood CR, Hecker M, Hosoya D, Hullo MF, Kakeshita H, Karamata D, Kasahara Y, Kawamura F, Koga K, Koski P, Kuwana R, Imamura D, Ishimaru M, Ishikawa S, Ishio I, Le Coq D, Masson A, Mauël C, Meima R, Mellado RP, Moir A, Moriya S, Nagakawa E, Nanamiya H, Nakai S, Nygaard P, Ogura M, Ohanan T, O'Reilly M, O'Rourke M, Pragai Z, Pooley HM, Rapoport G, Rawlins JP, Rivas LA, Rivolta C, Sadaie A, Sadaie Y, Sarvas M, Sato T, Saxild HH, Scanlan E, Schumann W, Seegers JFML, Sekiguchi J, Sekowska A, Séror SJ, Simon M, Stragier P, Studer R, Takamatsu H, Tanaka T, Takeuchi M, Thomaides HB, Vagner V, van Dijl JM, Watabe K, Wipat A, Yamamoto H, Yamamoto M, Yamamoto Y, Yamane K, Yata K, Yoshida K, Yoshikawa H, Zuber U and Ogasawara N (2003) Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci USA* **100**: 4678-4683

Korsak D, Markiewicz Z, Gutkind GO and Ayala JA (2010) Identification of the full set of *Listeria monocytogenes* penicillin-binding proteins and characterization of PBPD2 (Lmo2812) *BMC Microbiol* **10**: 239-252

Kostakioti M, Newman CL, Thanassi DG and Stathopoulos C (2005) Mechanisms of protein export across the bacterial outer membrane. *J Bacteriol* **187**: 4306-4314

Krissinel E and Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**: 774-797

Laupland KB, Parkins MD, Gregson DB, Church DL, Ross T and Pitout JDD (2008) Population-based laboratory surveillance for *Serratia* species isolates in a large Canadian health region. *Eur J Clin Microbiol Infect Dis* **27**: 89-95

Leach AR, Hann MM, Burrows JN and Griffen EJ (2006) Fragment screening: an introduction. *Mol BioSyst* **2**: 429-446

Leiman PG, Basler M, Ramagopal UA, Bonanno JB, Sauder JM, Pukatzki S, Burley SK, Almo SC and Mekalanos JJ (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci USA* **106**: 4154-4159

Lerner MG and Carlson HA (2006) APBS plugin for PyMOL. *Univeristy of Michagan*

Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, Villanueva J, Wei T and Ausubel FM (2006) An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc Natl Acad Sci USA* **103**: 2833-2838

Li M, Le Trong I, Carl M, Larson E, Chou S, De Leon J, Dove SL, Stenkamp RE and Mougous JD (2012) Structural basis for Type VI secretion effector recognition by a cognate immunity protein. *PLoS Pathog* **8**: e1002613

Lipinski C and Hopkins A (2004) Navigating chemical space for biology and medicine. *Nature* **432**: 855-861

Lipinski CA, Lombardo F, Dominy BW and Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46**: 3-26

Lovering AL, Safadi SS and Strynadka NCJ (2012) Structural perspective of peptidoglycan biosynthesis and assembly. *Annu Rev Biochem* **81**: 451-478

Ma AT, McAuley S, Pukatzki S and Mekalanos JJ (2009) Translocation of a *Vibrio cholerae* Type VI secretion effector requires bacterial endocytosis by host cells. *Cell Host Microbe* **5**: 234-243

Ma L-S, Narberhaus F and Lai E-M (2012) The IcmF-family protein TssM exhibits ATPase activity and energizes type VI secretion. *J Biol Chem* **287**: 15610-15621

Madeira A, Öhman E, Nilsson A, Sjögren B, Andrén PE and Svenningsson P (2009) Coupling surface plasmon resonance to mass spectrometry to discover novel protein-protein interactions. *Nat Protoc* **4**: 1023-1037

Magasanik B, Moyed H and Gehring L (1957) Enzymes essential for the biosynthesis of nucleic acid guanine; inosine 5'-phosphate dehydrogenase of *Aerobacter aerogenes*. *J Biol Chem* **226**: 339-350

Mahlen SD (2011) *Serratia* infections: from military experiments to current practice. *Clin Microbiol rev* **24**: 755-791

Maniatis T (1983) Molecular cloning. A laboratory manual by T Maniatis, E F Fritsch and J Sambrook. pp 545. Cold Spring Harbor Laboratory, New York. 1982. ISBN 0-87969-136-0. *Biochemical Education* **11**: 82-82

Matsumura M and Signor G, Matthews BW (1989) Substantial increase of protein stability by multiple disulphide bonds. *Nature* **342**: 291-293

Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33**: 491-497

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *J Appl Cryst* **40**: 658-674

McMillan FM, Cahoon M, White A, Hedstrom L, Petsko GA and Ringe D (2000) Crystal structure at 2.4 A resolution of *Borrelia burgdorferi* inosine 5'-monophosphate dehydrogenase: evidence of a substrate-induced hinged-lid motion by loop 6. *Biochemistry* **39**: 4533-4542

Miller RL and Adamczyk DL (1976) Inosine 5′-monophosphate dehydrogenase from Sarcoma 180 cells - substrate and inhibitor specificity. *Biochem Pharmacol* **25**: 883-888

Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, Goodman AL, Joachimiak G, Ordoñez CL, Lory S, Walz T, Joachimiak A and Mekalanos JJ (2006) A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* **312**: 1526-1530

Mueller CA, Broz P and Cornelis GR (2008) The type III secretion system tip complex and translocon. *Mol Microbiol* **68**: 1085-1095

Murdoch SL, Trunk K, English G, Fritsch MJ, Pourkarimi E and Coulthurst SJ (2011) The opportunistic pathogen *Serratia marcescens* utilizes Type VI secretion to target bacterial competitors. *J Bacteriol* **193**: 6057-6069

Murray CW and Blundell TL (2010) Structural biology in fragment-based drug design. *Curr Opin Struct Biol* **20**: 1-11

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F and Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst* D**67**: 355-367

Nagano N, Orengo CA and Thornton JM (2002) One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *Technology (AIST)* **2**: 41-46

Neumann T, Junker H-D, Schmidt K and Sekul R (2007) SPR-based fragment screening: advantages and applications. *Curr Top in Med Chem* **7**: 1630-1642

Niesen FH, Berglund H and Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**: 2212-2221

Nimmesgern E, Black J, Futer O, Fulghum JR, Chambers SP, Brummel CL, Raybuck SA and Sintchak MD (1999) Biochemical analysis of the modular enzyme inosine 5'-monophosphate dehydrogenase. *Protein expression and purification* **17**: 282-289

Obritsch MD, Fish DN, MacLaren R and Jung R (2004) National surveillance of antimicrobial resistance in *Pseudomonas aeruginosa* isolates obtained from intensive care unit patients from 1993 to 2002. *Antimicrob Agents Chemother* **48**: 4606-4610

Painter J and Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. *J Appl Cryst* **39**: 109-111

Pantoliano M, Petrella E and Kwasnoski J (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **6**: 429-440

Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR and Schacht AL (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* **9**: 203-214

Peleg AY and Hooper DC (2010) Hospital-acquired infections due to Gram-negative bacteria. *N Engl J Med* **362**: 1804-1813

Pell LG, Kanelis V, Donaldson LW, Howell PL and Davidson AR (2009) The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc Natl Acad Sci USA* **106**: 4160-4165

Pellecchia M, Bertini I, Cowburn D, Dalvit C, Giralt E, Jahnke W, James TL, Homans SW, Kessler H, Luchinat C, Meyer B, Oschkinat H, Peng J, Schwalbe H and Siegal G (2008) Perspectives on NMR in drug discovery: a technique comes of age. *Nat Rev Drug Discov* **7**: 738-745

Petersen TN, Brunak S, von Heijne G and Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth* **8**: 785-786

Pietrosiuk A, Lenherr ED, Falk S, Boenemann G, Kopp J, Zentgraf H, Sinning I and Mogk A (2011) Molecular basis for the unique role of the AAA+ chaperone ClpV in type VI protein secretion. *J Biol Chem* **286**: 30010-30021

Pimkin M and Markham GD (2008) The CBS subdomain of inosine 5'-monophosphate dehydrogenase regulates purine nucleotide turnover. *Mol Microbiol* **68**: 342-359

Powell H, Leslie A, and Battye G (2007) Mosflm 7.0.1 and its new interface - iMosflm 0.5.3 *CCP4 Newsletter* **46**

Prapunpoj P and Leelawatwattana L (2009) Evolutionary changes to transthyretin: structure–function relationships. *FEBS J* **276**: 5330-5341

Pukatzki S, Ma AT, Revel AT, Sturtevant D and Mekalanos JJ (2007) Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc Natl Acad Sci USA* **104**: 15508-15513

Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, Heidelberg JF and Mekalanos JJ (2006) Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc Natl Acad Sci USA* **103**: 1528-1533

Rao VA, Shepherd SM, English G, Coulthurst SJ and Hunter WN (2011) The structure of *Serratia marcescens* Lip, a membrane-bound component of the type VI secretion system. *Acta Cryst* D**67**: 1065-1072

Records AR (2011) The type VI secretion system: a multipurpose delivery system with a phage-like machinery. *Mol Plant Microbe Interact* **24**: 751-757

Rees DC, Congreve M, Murray CW and Carr R (2004) Fragment-based lead discovery. *Nat Rev Drug Discov* **3**: 660-672

Rossmann MG, Blow DM (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst* **15**: 24-31

Rupp B (2010) Biomolecular crystallography: principles, practice, and application to structural biology ISBN 978-0-8153-4081-2

Russell AB, Hood RD, Bui NK, Leroux M, Vollmer W and Mougous JD (2011) Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **475**: 343-347

Sainsbury S, Bird L, Rao V, Shepherd SM, Stuart DI, Hunter WN, Owens RJ and Ren J (2011) Crystal structures of penicillin-binding protein 3 from *Pseudomonas aeruginosa*: comparison of native and antibiotic-bound forms. *J Mol Biol* **405**: 173-184

Sauvage E, Kerff F, Terrak M, Ayala JA and Charlier P (2008) The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol Rev* **32**: 234-258

Sana TG, Hachani A, Bucior I, Soscia C, Garvis S, Termine E, Engel J, Filloux A and Bleves S (2012) The second type six secretion system of *Pseudomonas aeruginosa* strain PAO1 is regulated by quorum sensing and Fur and modulates internalization in epithelial cells. *J Biol Chem* epub ahead of print

Schneider T and Sahl H-G (2010) An oldie but a goodie - cell wall biosynthesis as antibiotic target pathway. *Int J Med Microbiol* **300**: 161-169

Schrodinger LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3r1

Scott DE, Coyne AG, Hudson SA and Abell C (2012) Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **51**: 4990-5003

Scott R (2009) The direct medical costs of healthcare-associated infections in U.S. hospitals and the benefits of prevention. *Centres for disease control and prevention*: http://www.cdc.gov/HAI/pdfs/hai/Scott_CostPaper.pdf

Shahian T, Lee GM, Lazic A, Arnold LA, Velusamy P, Roels CM, Guy RK and Craik CS (2009) Inhibition of a viral enzyme by a small-molecule dimer disruptor. *Nat Chem Biol* **5**: 640-646

Shalom G, Shaw JG and Thomas MS (2007) *In vivo* expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. *Microbiol* **153**: 2689-2699

Shouldice SR, Heras B, Walden PM, Totsika M, Schembri MA and Martin JL (2011) Structure and function of DsbA, a key bacterial oxidative folding catalyst. *Antioxid Redox Signal* **14**: 1729-1760

Sintchak MD and Nimmesgern E (2000) The structure of inosine 5'-monophosphate dehydrogenase and the design of novel inhibitors. *Immunopharmacology* **47**:163-184

Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA and Godzik A (2007) XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**: 3403-3405

Sleigh JD (1983) Antibiotic resistance in *Serratia marcescens*. *BMJ* **287**: 1651-1653

Souli M, Galani I and Giamarellou H (2008) Emergence of extensively drug-resistant and pandrug-resistant Gram-negative bacilli in Europe. *Euro Surveill* **13**: 1-11

Spratt BG (1975) Distinct penicillin binding proteins involved in the division, elongation, and shape of *Escherichia coli* K12. *Proc Natl Acad Sci USA* **72**: 2999-3003

Spratt BG (1977) Properties of the penicillin-binding proteins of *Escherichia coli* K12. *Eur J Biochem* **72**: 341-352

Spratt BG and Cromie KD (1988) Penicillin-binding proteins of Gram-negative bacteria. *Rev Infect Dis* **10**: 699-711

Sprenger M (2010) Annual epidemiological report on communicable diseases in Europe 2010. *European Centre for Disease Prevention and Control* ISBN 978-92-9193-222-1

Studier F (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* **41**: 207-234

Talbot GH, Bradley J, Edwards JE, Gilbert D, Scheld M and Bartlett JG (2006) Bad bugs need drugs: an update on the development pipeline from the Antimicrobial Availability Task Force of the Infectious Diseases Society of America. *Clin Infect Dis* **42**: 657-668

Taylor GL (2010) Introduction to phasing. *Acta Cryst* D**66**: 325-338

Terwilliger TC (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Cryst* D**59**: 38-44

Thanabalu T, Koronakis E, Hughes C and Koronakis V (1998) Substrate-induced assembly of a contiguous channel for protein export from *E.coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *EMBO J* **17**: 6487-6496

Tokuda H and Matsuyama S-I (2004) Sorting of lipoproteins to the outer membrane in *E. coli*. *Biochim Biophys Acta* **1693**: 5-13

Typas A, Banzhaf M, Gross CA and Vollmer W (2011) From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat Rev Micro* **10**: 123-136

Umejiego NN, Gollapalli D, Sharling L, Volftsun A, Lu J, Benjamin NN, Stroupe AH, Riera TV, Striepen B and Hedstrom L (2008) Targeting a prokaryotic protein in a eukaryotic pathogen: identification of lead compounds against cryptosporidiosis. *Chem Biol* **15**: 70-77

van Heijenoort J (2001) Formation of the glycan chains in the synthesis of bacterial peptidoglycan. *Glycobiology* **11**: 25R-36R

van Heijenoort J and Gutmann L (2000) Correlation between the structure of the bacterial peptidoglycan monomer unit, the specificity of transpeptidation and susceptibility to beta-lactams. *Proc Natl Acad Sci USA* **97**: 5028-5030

Van Rheenen SM, Duménil G and Isberg RR (2004) IcmF and DotU are required for optimal effector translocation and trafficking of the *Legionella pneumophila* vacuole. *Infect Immun* **72**: 5972-5982

Vicente M, Hodgson J, Massidda O, Tonjum T, Henriques-Normark B and Ron EZ (2006) The fallacies of hope: will we discover new antibiotics to combat pathogenic bacteria in time? *FEMS Microbiol Rev* **30**: 841-852

Vollmer W and Bertsche U (2008) Murein (peptidoglycan) structure, architecture and biosynthesis in *Escherichia coli*. *Biochim Biophys Acta* **1778**: 1714-1734

Vollmer W, Blanot D and de Pedro MA (2008) Peptidoglycan structure and architecture. *FEMS Microbiol Rev* **32**: 149-167

Warrick JW, Murphy WL and Beebe DJ (2008) Screening the cellular microenvironment: a role for microfluidics. *Biomedical Engineering, IEEE Rev* **1**: 75-93

Wartchow CA, Podlaski F, Li S, Rowan K, Zhang X, Mark D and Huang K-S (2011) Biosensor-based small molecule fragment screening with biolayer interferometry. *J Comput Aided Mol Des* **25**: 669-676

Weber B, Hasic M, Chen C, Wai SN and Milton DL (2009) Type VI secretion modulates quorum sensing and stress response in *Vibrio anguillarum*. *Environ Microbiol* **11**: 3018-3028

Weber G (1983) Biochemical strategy of cancer cells and the design of chemotherapy: G. H. A. Clowes Memorial Lecture. *Cancer Res* **43**: 3466-3492

Wetzel R, Perry LJ, Baase WA and Becktel WJ (1988) Disulfide bonds and thermal stability in T4 lysozyme. *Proc Natl Acad Sci USA* **85**: 401-405

Wheat R, Zuckerman A and Rantz L (1951) Infection due to chromobacteria; report of 11 cases. *AMA Arch Intern Med* **88**: 461-466

Whitby FG, Luecke H, Kuhn P, Somoza JR, Huete-Perez JA, Phillips JD, Hill CP, Fletterick RJ and Wang CC (1997) Crystal structure of *Tritrichomonas foetus* inosine-5'-monophosphate dehydrogenase and the enzyme-product complex. *Biochemistry* **36**: 10666-10674

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A and Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Cryst* D**67**: 235-242

Yamaguchi Y and Inouye M (2011) Regulation of growth and death in *Escherichia coli* by toxin–antitoxin systems. *Nat Rev Micro* **9**: 779-790

Zhang R, Evans G, Rotella F, Westbrook E, Huberman E, Joachimiak A and Collart F (1999a) Differential signatures of bacterial and mammalian IMP dehydrogenase enzymes. *Curr Med Chem* **6**: 537-543

Zhang R, Evans G, Rotella FJ, Westbrook EM, Beno D, Huberman E, Joachimiak A and Collart FR (1999b) Characteristics and crystal structure of bacterial inosine-5'-monophosphate dehydrogenase. *Biochemistry* **38** : 4691-700

Zheng J, Ho B and Mekalanos JJ (2011) Genetic analysis of anti-amoebae and anti-bacterial activities of the type VI secretion system in *Vibrio cholerae*. *PLoS ONE* **6**: e23876

Zheng J and Leung KY (2007) Dissection of a type VI secretion system in *Edwardsiella tarda*. *Mol Microbiol* **66**: 1192-1206

Zou TT, Yao X, Qin B, Zhang M, Cai LF, Shang W, Svergun DI, Wang M, Cui S and Jin Q (2012) Crystal structure of *Pseudomonas aeruginosa* Tsi2 reveals a stably folded superhelical antitoxin. *J Mol Biol* **417**: 351-361

APPENDICES

APPENDIX A

Sainsbury S, Bird L, Rao VA, Shepherd SM, Stuart DI, Hunter WN, Owens RJ and Ren J (2011) Crystal structures of penicillin-binding protein 3 from *Pseudomonas aeruginosa*: comparison of native and antibiotic-bound forms. ***Journal of Molecular Biology*** **405: 173-184**

doi:10.1016/j.jmb.2010.10.024 J. Mol. Biol. (2011) 405, 173–184

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Crystal Structures of Penicillin-Binding Protein 3 from *Pseudomonas aeruginosa*: Comparison of Native and Antibiotic-Bound Forms

Sarah Sainsbury[1], Louise Bird[1,2], Vincenzo Rao[3], Sharon M. Shepherd[3], David I. Stuart[1], William N. Hunter[3], Raymond J. Owens[1,2]* and Jingshan Ren[1]*

[1]*Division of Structural Biology, Henry Wellcome Building for Genomic Medicine, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK*
[2]*Oxford Protein Production Facility UK, The Research Complex at Harwell, Rutherford Appleton Laboratory Harwell Science and Innovation Campus, Oxfordshire OX11 0FA, UK*
[3]*Biological Chemistry and Drug Discovery, College of Life Sciences, The Wellcome Trust Building, University of Dundee, Dow Street, Dundee DD1 5EH, UK*

Received 16 August 2010; received in revised form 8 October 2010; accepted 15 October 2010
Available online 23 October 2010

Edited by I. Wilson

Keywords:
anti-bacterial;
Pseudomonas aeruginosa;
carbenicillin;
ceftazidime;
enzyme structure

We report the first crystal structures of a penicillin-binding protein (PBP), PBP3, from *Pseudomonas aeruginosa* in native form and covalently linked to two important β-lactam antibiotics, carbenicillin and ceftazidime. Overall, the structures of apo and acyl complexes are very similar; however, variations in the orientation of the amino-terminal membrane-proximal domain relative to that of the carboxy-terminal transpeptidase domain indicate interdomain flexibility. Binding of either carbenicillin or ceftazidime to purified PBP3 increases the thermostability of the enzyme significantly and is associated with local conformational changes, which lead to a narrowing of the substrate-binding cleft. The orientations of the two β-lactams in the active site and the key interactions formed between the ligands and PBP3 are similar despite differences in the two drugs, indicating a degree of flexibility in the binding site. The conserved binding mode of β-lactam-based inhibitors appears to extend to other PBPs, as suggested by a comparison of the PBP3/ceftazidime complex and the *Escherichia coli* PBP1b/ceftoxamine complex. Since *P. aeruginosa* is an important human pathogen, the structural data reveal the mode of action of the frontline antibiotic ceftazidime at the molecular level. Improved drugs to combat infections by *P. aeruginosa* and related Gram-negative bacteria are sought and our study provides templates to assist that process and allows us to discuss new ways of inhibiting PBPs.

**Corresponding authors.* E-mail addresses: raymond.owens@rc-harwell.ac.uk; ren@strubi.ox.ac.uk.

Present address: S. Sainsbury, Gene Center Munich, University of Munich—LMU, Feodor-Lynen Strasse 25, 81377 Munich, Germany.

Abbreviations used: PBP, penicillin-binding protein; HMM, high molecular mass; LMM, low molecular mass; PDB, Protein Data Bank; ESRF, European Synchrotron Radiation Facility.

Introduction

The bacterial cell wall consists of *N*-acetylglucosamine-*N*-acetyl muramic polymers cross-linked *via* the penultimate D-alanine residues of muramyl pentapeptides to form a peptidoglycan network that is essential for normal cell survival. Penicillin-binding proteins (PBPs) are membrane-bound enzymes involved in the final stages of bacterial cell wall synthesis on the periplasmic side of the membrane. They have been classified into a high-molecular-mass (HMM) group, members of which are essential for cell viability, and a low-molecular-mass (LMM) group, members of which appear dispensable for normal cell growth.[1] HMM PBPs are further divided into class A enzymes, which catalyze both the polymerization of a peptidoglycan from disaccharide peptides (glycosyltransferase) and the cross-linking of muramyl peptides (transpeptidase), and class B enzymes, which only possess transpeptidase activity. HMM PBPs consist of two domains: an N-terminal membrane-proximal domain and a C-terminal domain harboring transpeptidase activity. In the case of class A PBPs, the N-terminal domain contains glycosyltransferase activity; the function of the equivalent domain in class B PBPs is not known. The smaller LMM PBPs are D,D-carboxypeptidases that remove the terminal D-alanine from the muramyl peptide, thus controlling the level of peptidoglycan cross-linking.[1] As the name indicates, PBPs are also the primary targets of β-lactam antibiotics, which act as suicide substrates by mimicking the D-alanyl-D-alanine stem peptide of peptidoglycan precursors. β-Lactams block the activity of transpeptidases and carboxypeptidases by acylating the active site serine residue.[2]

Crystal structures have been reported for HMM PBPs from the Gram-negative *Escherichia coli* (PBP1b)[3] and *Neisseria gonorrhoeae* (PBP2),[4] and the Gram-positive *Streptococcus pneumoniae* (PBP1b,[5] PBP2x,[6–10] and PBP2b[11]) and *Staphylococcus aureus* (PBP2a[12]), including both apo-enzymes and complexes with β-lactams. All of these structures show similar active sites, featuring three conserved catalytic motifs in the transpeptidase domain. The β-lactams are covalently bound to the active site serine *via* the carbonyl of the opened cyclic amide. Resistance to β-lactams can be rationalized to some extent by examining the structures of PBPs from resistant bacteria, which have mutations in residues close to the catalytic motifs.[2,7,11]

Our interest is primarily *Pseudomonas aeruginosa*, an opportunistic pathogen that is responsible for severe life-threatening infections, especially in patients suffering from chronic respiratory diseases such as cystic fibrosis.[13] The genome of *P. aeruginosa* strain PAO1[14] encodes four HMM PBPs, including one class A enzyme (PBP1a) and three class B enzymes (PBP2, PBP3, and PBP3a), which are orthologues of the corresponding enzymes in *E. coli*. PBP3 shares 42% sequence identity with the corresponding protein in *E. coli* and has been identified as the primary target of a number of β-lactams used to treat pseudomonal infections, including the cephalosporin analogues cefsulodin[15] and ceftazidime,[16] piperacillin,[17] and the parenteral carbapenem, doripenem.[18]

Compared to many other Gram-negative bacteria, *P. aeruginosa* is highly resistant to antibiotics and, as with many bacterial pathogens, resistance increases with repeated use and misuse of antibiotics. As is typical of Gram-negative bacteria, antibiotic resistance in *P. aeruginosa* is largely due to a combination of β-lactamase production and the action of numerous efflux pumps.[19] However, treatment with high doses of piperacillin has been shown to be associated with reduced binding to PBPs in clinical isolates of *P. aeruginosa*.[17] In laboratory strains lacking β-lactamases, the overproduction of PBP3 in *P. aeruginosa* promotes reduced sensitivity to cefsulodin;[20] in a separate study, increased resistance to cefsulodin was observed to be associated with reduced binding to endogenous PBP3.[15]

PBP3 is a therapeutic target in *P. aeruginosa*, and we have determined structures of this HMM PBP in its native form and in complex with two relevant drugs, carbenicillin and ceftazidime (Supplementary Material, Fig. S1). Carbenicillin is a broad-spectrum semisynthetic penicillin derivative and was one of the first derivatives to display significant activity against *Pseudomonas*.[21] The aminothiazolyl-carrying ceftazidime[16] is similar to cefotaxime, the first third-generation cephalosporin to be marketed, but differs with a 2-carboxy-2-oxypropaneimino group replacing a methoxy amino group. This difference increases efficacy against *P. aeruginosa*, making ceftazidime one of only a handful of drugs currently used to treat *Pseudomonas* infections. Ceftazidime also displays potent activity against another of the so-called superbugs, *Burkholderia pseudomallei*, which was previously known as *Pseudomonas pseudomallei* and is the causal agent of melioidosis, a serious infection in humans found mainly in Asia. Indeed, ceftazidime is the frontline drug for the treatment of this disease.

We describe the crystal structures of apo-PBP3 and the acyl complexes with carbenicillin and ceftazidime, as well as the similarities and differences between the three structures, and discuss structural features that might be exploited to derive novel inhibitors of therapeutic value.

Results and Discussion

Production of PBP3 and binding to β-lactams

A soluble version of recombinant *P. aeruginosa* PBP3 was produced by truncating the sequence at the N-terminus by 34 residues to remove the

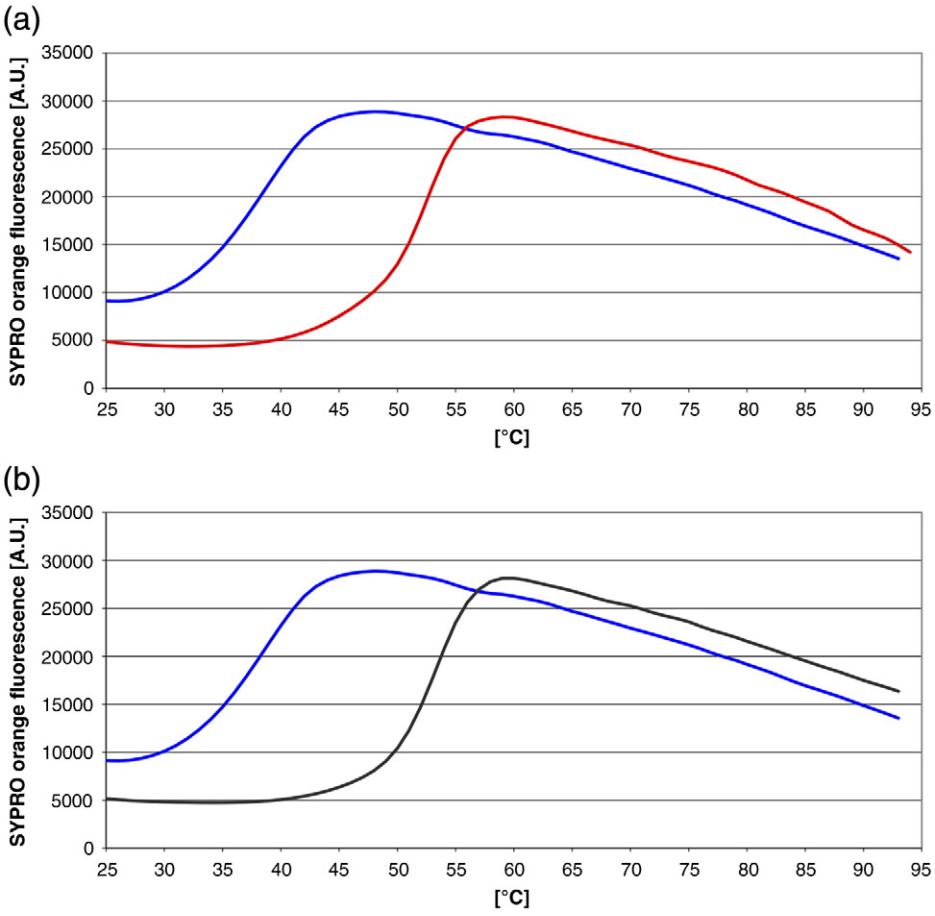


Fig. 1. Thermal shift assay. (a) PBP3 screened with 500 μM carbenicillin, giving a ΔT_m of +13.2 °C. The blue curve is the native protein, and the red curve is the complex. (b) PBP3 screened with 500 μM ceftazidime, giving a ΔT_m of +14.5 °C. The blue curve is the native protein, and the black curve is the complex.

predicted transmembrane helix. The molecular mass of the purified protein determined by size-exclusion chromatography was approximately 54 kDa compared to a calculated value of 61 kDa, indicating that it is most likely a monomer in solution. Binding of carbenicillin and ceftazidime to the recombinant protein was confirmed using a thermal shift assay. The assay used a real-time PCR machine to detect changes in fluorescence during the thermal unfolding of the protein in the presence of SYPRO Orange. Significant temperature stabilization of the protein was observed for both compounds, with shifts in the midpoint transition temperature (ΔT_m) of 13.2 ± 0.2 °C and 14.5 ± 0.1 °C following incubation with either 0.5 mM carbenicillin or 0.5 mM ceftazidime, respectively (Fig. 1).

Overall structure

Three isomorphous structures of *P. aeruginosa* PBP3 have been determined: apo-PBP3 at 2.0 Å resolution, PBP3/carbenicillin complex at 2.4 Å resolution, and PBP3/ceftazidime complex at 2.6 Å resolution. The PBP3/carbenicillin complex was derived from a sample obtained by cocrystallization, and the ceftazidime complex was derived by soaking a native crystal in mother liquor containing the antibiotic. The initial structure was determined using the coordinates of PBP2 from *N. gonorrhoeae*[4] as molecular replacement model [Protein Data Bank (PDB) code 3equ]. Crystallographic statistics (Table 1) indicate that acceptable models have resulted from the analyses. In all three structures, residues 35–49, 492–497, and 560–577 are disordered and do not have sufficient electron density to allow reliable model building.

PBP3 is a two-domain protein comprising a C-terminal transpeptidase linked to an extended N-terminal domain. Overall, PBP3 resembles other class B PBPs, although it lacks the additional C-terminal domain unique to the PBP2x structure (Fig. 2). In common with these other class B PBPs, there is an α-helical subdomain or "head" domain[11] (residues 80–149) towards the N-terminus, which

Table 1. X-ray data collection and refinement statistics

Data collection			
X-ray source	Diamond I04	ESRF ID23-2	Diamond I02
Data set	PBP3 native	PBP3/carbenicillin	PBP3/ceftazidime
Wavelength (Å)	1.0000	0.8726	0.9750
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters *a*, *b*, *c* (Å)	66.76, 81.51, 85.94	69.48, 83.44, 89.85	67.08, 81.81, 87.79
Resolution range (Å)	30.0–2.00 (2.07–2.00)	30.0–2.40 (2.49–2.40)	30.0–2.60 (2.69–2.60)
Unique reflections	32,803 (3150)	21,168 (2081)	15,047 (1331)
Completeness (%)	99.6 (96.7)	99.6 (99.9)	98.6 (89.3)
Redundancy	10.4 (5.7)	5.4 (4.9)	5.6 (3.7)
$\langle I/\sigma(I)\rangle$	23.9 (3.0)	8.6 (1.9)	11.9 (2.5)
R_{merge}	0.143 (0.575)	0.172 (0.582)	0.166 (0.545)
Refinement statistics			
Resolution range (Å)	30.0–2.00	30.0–2.40	30.0–2.60
Number of reflections (working/test)	31,100/1649	19,881/1090	14,178/764
R-factor[a] (R_{work}/R_{free})	0.171/0.213	0.195/0.242	0.207/0.267
Number of atoms (protein/other)	3787/220	3801/158	3841/69
RMSD on bond lengths (Å)	0.007	0.007	0.006
RMSD on bond angles (°)	1.1	1.1	1.0
Mean *B*-factor (protein/other) (Å^2)	32/38	30/33	21/16
Ramachandran plot			
Residues in preferred regions (%)	469 (95.7)	474 (96.9)	373 (88.2)
Residues in allowed regions (%)	21 (4.3)	14 (2.9)	46 (10.9)
Residues in disallowed regions (%)	0 (0.0)	1 (0.2)	1 (0.2)

[a] R_{work} and R_{free} are defined by $R = \sum_{hkl}||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$, where *hkl* are the indices of the reflections (used in refinement for R_{work}; 5%, not used in refinement, for R_{free}), and F_{obs} and F_{calc} are the structure factors deduced from measured intensities and calculated from the model, respectively.

was not resolved in the PBP2 structure (Fig. 2 and Table 2). The function of the N-terminal domain of class B PBPs is not known but possibly serves to position the transpeptidase domain away from the inner membrane as part of a multienzyme complex involved in cell wall biosynthesis.[24]

Despite a relatively low sequence identity, the fold of the transpeptidase domain is highly conserved among PBPs. For the class B enzymes, the RMSD for the overlap of approximately 80% of the C^{α} atoms is 1.5–2.0 Å, depending on the structure (Table 2). The structural superimpositions also revealed that the orientation of the transpeptidase domains varied with respect to the N-terminal domains in the different structures. For example, in the overlay of the PBP3 and PBP2b transpeptidase domains, the N-terminal domains of the PBPs were rotated as rigid bodies by about 24° with respect to each other (Table 2), indicating considerable rotational freedom between the N-terminal domain and the C-terminal domain.

Active site structure

The active site of PBP3 is located in a long cleft running parallel with the β3 strand across the lower part of the transpeptidase domain. The three conserved structural motifs that are common to D, D-peptidases and β-lactamases are found here and have well-defined electron density (Fig. 3a). The SXXK motif (S294-T295-V296-K297) is located at the N-terminal end of α2 and includes the nucleophilic S294, together with the α8–β2f loop (residues 407–409) that forms the base of the cleft. The SXN motif (S349-S350-N351) occurs in the α4–α5 turn, and residues 328–334 (β2c and β2c–β2d loop) form one side of the elongated cleft; the third motif KS/TG (K484-S485-G486) is positioned on strand β3 and, together with residues 530–533 in the β5–α11 loop, forms the opposite side of the cleft. Unlike in PBP2a,[12] which is relatively resistant to β-lactam inhibition, the catalytic S294 is positioned favorably for acylation in an open substrate-binding pocket. The three conserved structural motifs interact with each other *via* a network of hydrogen bonds: the hydroxyl group of S294 forms hydrogen bonds with the side chains of both K297 and K484 (2.7 Å and 2.8 Å), which are hydrogen bonded to S349 and N351, respectively (Fig. 3d). The position of K297 makes it one of the possible proton receptors from the catalytic S294 during the acylation reaction, in addition to the possible contribution of the β-lactam carboxylate. There are eight water molecules in the active site cleft of apo-PBP3, which form an extensive hydrogen-bonding network with the protein. The two loops connecting β3 and β4 (K490-A500) and β5 and α11 (S526-F531) are disordered in the apo-enzyme.

Comparison of apo-enzyme and acyl-enzyme complexes

In comparing the apo-PBP3 and acyl-PBP3 complexes, it is apparent that the N-terminal domains

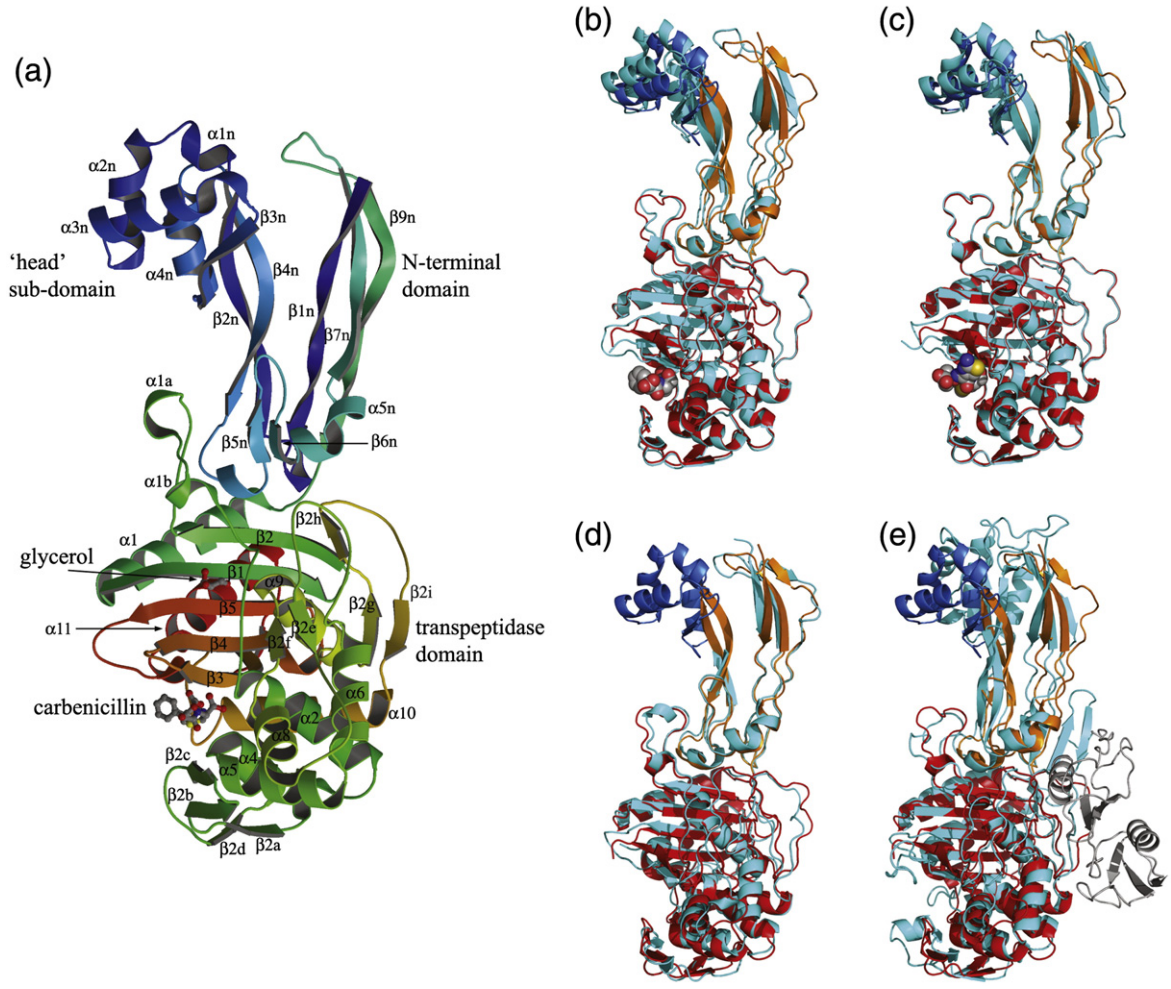


Fig. 2. The overall structure of PBP3. (a) A ribbon diagram of the PBP3/carbenicillin complex showing the overall fold, antibiotic binding site, and bound glycerol. The diagram is rainbow-colored from blue at the N-terminus to red at the C-terminus, and the secondary structure elements are labelled in accordance with previous studies.[15,16] (b–e) A comparison of the overall structure of native PBP3 (blue, orange, and red) with those of the PBP3/carbenicillin complex (b), PBP3/ceftazidime complex (c), PBP2 of *N. gonorrhoeae* (d), and PBP2x from *S. pneumoniae* (e). This figure was produced with PyMOL (www.pymol.org). The rest of the figures were drawn using BOBSCRIPT.[22]

show greater structural flexibility than the C-terminal transpeptidase domains. Thus, the superposition of the N-terminal β subdomains (residues 50–79 and 150–221) of the PBP3/carbenicillin and PBP3/ceftazidime complexes with the apo-enzyme resulted in RMSDs of 1.3 Å and 1.1 Å for 166 and 169 equivalent C^{α} atoms, and 7° and 5° differences in orientation for the head subdomains, respectively. By contrast, the C-terminal transpeptidase domains of both acyl-PBP3 enzymes superimpose on the apo-enzyme with RMSDs of 0.8 Å and 0.6 Å for the carbenicillin complex and the ceftazidime complex for 319 of 324 overlapped C^{α} positions. This shows that acylation does not significantly alter the overall conformation of the transpeptidase domain. However, significant local changes are observed in the substrate-binding pocket with movement of β3 relative to α2, reinstating the strict anti-parallel arrangement of the three β-strands (β3, β4, and β5). Strands β3 and β4 of PBP3 bend towards the inhibitor around a hinge region located at residue G486 on β3 and at residue L504 on β4, resulting in an approximately 4-Å movement for the C^{α} atoms of K490 and A500, the last and the first ordered residues of the β3 and β4 strands in the apo structure, respectively (Fig. 4a). A similar movement of the equivalent strands has been observed following the binding of β-lactams to other HMM[5] and LMM PBPs.[25] Ligand binding is associated with an increased order of the β3–β4 and β5–α11 loops, as shown by better electron density for these residues in the two complexes. Both carbenicillin and ceftazidime display well-defined electron density in the acylated enzyme complexes (Fig. 3b and c). The β-lactams are covalently bound

Table 2. A comparison of class B PBP structures

PDB code	Number of C^{α} equivalences	Identical sequence	RMSD (Å)	Rotation (Å)	Translation (Å)
Overlap of the C-domain onto PBP3 native (residues 222–558)					
2wad-A	267	59	1.7	NA	
1rp5-A	293	75	1.9	NA	
1mwr-A	262	57	2.0	NA	
3equ-A	302	126	1.5	NA	
Overlap of the N-terminal subdomain (residues 50–79 and 150–221)					
2wad-A	84	20	1.8	24.9	1.1
1rp5-A	97	28	1.3	2.5	0.4
1mwr-A	71	24	1.4	17.0	1.0
3equ-A	90	41	1.7	8.9	0.1
Overlap of the head subdomain (residues 80–149)					
2wad-A	60	8	2.5	28.7	1.0
1rp5-A	67	7	1.9	60.3	2.1
1mwr-A	60	11	2.1	76.6	0.3
3equ-A					

NA, not applicable.

The structures of the C-terminal transpeptidase domain, N-terminal β-stranded subdomain, and N-terminal β-stranded head subdomain were sequentially aligned to PBP3 using SHP.[23] The results show the RMSD for C^{α} atoms, and the rotation and translation distances required to overlay all of the domains. In each case, molecule A in the PDB entries for the following structures was used for the comparison: *S. pneumoniae* PBP2b (PDB code 2wad), *S. pneumonia* PBP2x (PDB code 1rp5), *S. aureus* PBP2a (PDB code 1mwr), and *N. gonorrhoeae* PBP2 (PDB code 3equ).

to the active site S294 and adopt similar orientations that are maintained by hydrogen-bonding interactions between S485 and T487 and a carboxylate group on both compounds (Fig. 3d and e). The alignment of the drugs with T487 resembles a short segment of anti-parallel β-sheet. Note that ceftazidime is susceptible to hydrolysis,[26] and since the R_2 piperidine moiety of ceftazidime is lost on β-lactam ring opening, both compounds have similar chemical groups at this position [i.e., one methyl (ceftazidime) or two methyls (carbenicillin)], and these are positioned near V333 and F531 (Fig. 3g). The principal difference between the inhibitors involves the nature and the size of the R_1 substituent (Fig. 3b and c; Supplementary Material, Fig. S1). In both complexes, the R_1 amide linkage forms hydrogen bonds with the main-chain carbonyl of T487 and with the side chain of N351, and a cluster of aromatic residues (Y501, Y530, and F531) forms hydrophobic contacts with the R_1 groups (Fig. 3e and f). In the carbenicillin complex, the R_1 carboxylate forms hydrogen bonds with the side-chain hydroxyl of Y409 and a water molecule, which in turn interacts with the main-chain amide of R489. In order to accommodate the much larger R_1 substituent on ceftazidime, the side-chain of Y409 is rotated away from the position observed in the apo-structure. This movement creates a subpocket where van der Waals interactions involving E291, G293, and A488 help to position the R_1 aminothiazole ring and allows ring-stacking interactions with Y409. The position of the aminothiazole ring is also stabilized through bifurcated hydrogen bonds to the main-chain carbonyl of R489 and to the side-chain of E291, which is the only negatively charged residue in the cleft (Fig. 3g). The main-chain amide of R489 donates a hydrogen bond to the thiazole nitrogen. In the ceftazidime complex, the side-chain of R489 is rotated towards the ligand, allowing the side chain to donate a hydrogen bond to the R_1 dimethyl/carboxylate group (Fig. 3f and g).

Conservation of active-site residues in *B. pseudomallei* PBP3

As mentioned earlier, ceftazidime is used to treat infections with *B. pseudomallei*. The amino acid sequences of PBP3 from *P. aeruginosa* and *B. pseudomallei* were aligned (accession numbers Q51504 and Q63QJ1, respectively; data not shown). The enzymes share an approximately 40% identity. Of 12 residues (in and around the active site) that are critical for *P. aeruginosa* PBP3 reactivity and interaction with ceftazidime (K297, V333, S349, N351, Y409, K484, S485, T487, R489, Y501, Y530, and F531), 9 are strictly conserved. The differences are that V333 changes to aspartate, R489 changes to tyrosine, and Y530 changes to histidine. These are sterically conservative changes (e.g., the aliphatic component of the aspartate side-chain matches closely to a part of valine) and suggest that our model of the PBP3/ceftazidime complex reveals the structural basis for the action of ceftazidime in the treatment of melioidosis. Furthermore, this model provides a template for informing inhibitor development.

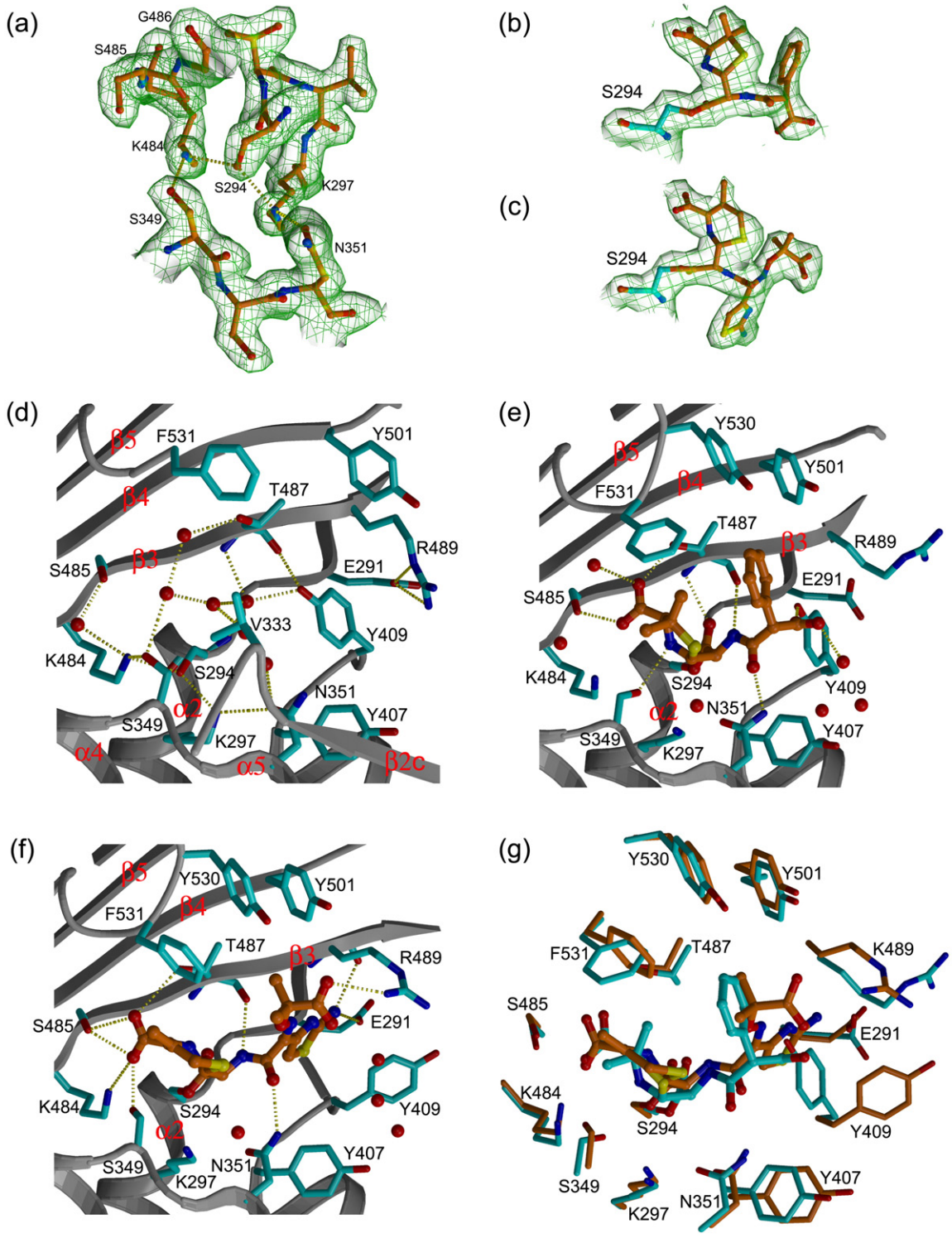


Fig. 3. The active site of PBP3. (a–c) $2F_o - F_c$ maps calculated after the final round of refinement and contoured at 1σ showing the electron density for the three active site structural motifs in apo-PBP3, bound carbenicillin, and ceftazidime in the two complexes, respectively. In (b) and (c), the carbon atoms of the acylated S294 are shown in cyan. (d–f) The active sites of apo-PBP3, carbenicillin, and ceftazidime complexes. The protein main chains are shown as gray ribbons, side chains are drawn as cyan sticks, and the inhibitor is shown in orange ball-and-sticks. The red spheres represent water molecules, and the yellow broken lines represent potential hydrogen bonds. Residue V333 that makes contacts with the bound inhibitors is not shown in (e) and (f) for clarity. (g) Comparison of the active sites of the PBP3/carbenicillin complex (cyan) and the PBP3/ceftazidime complex (orange).

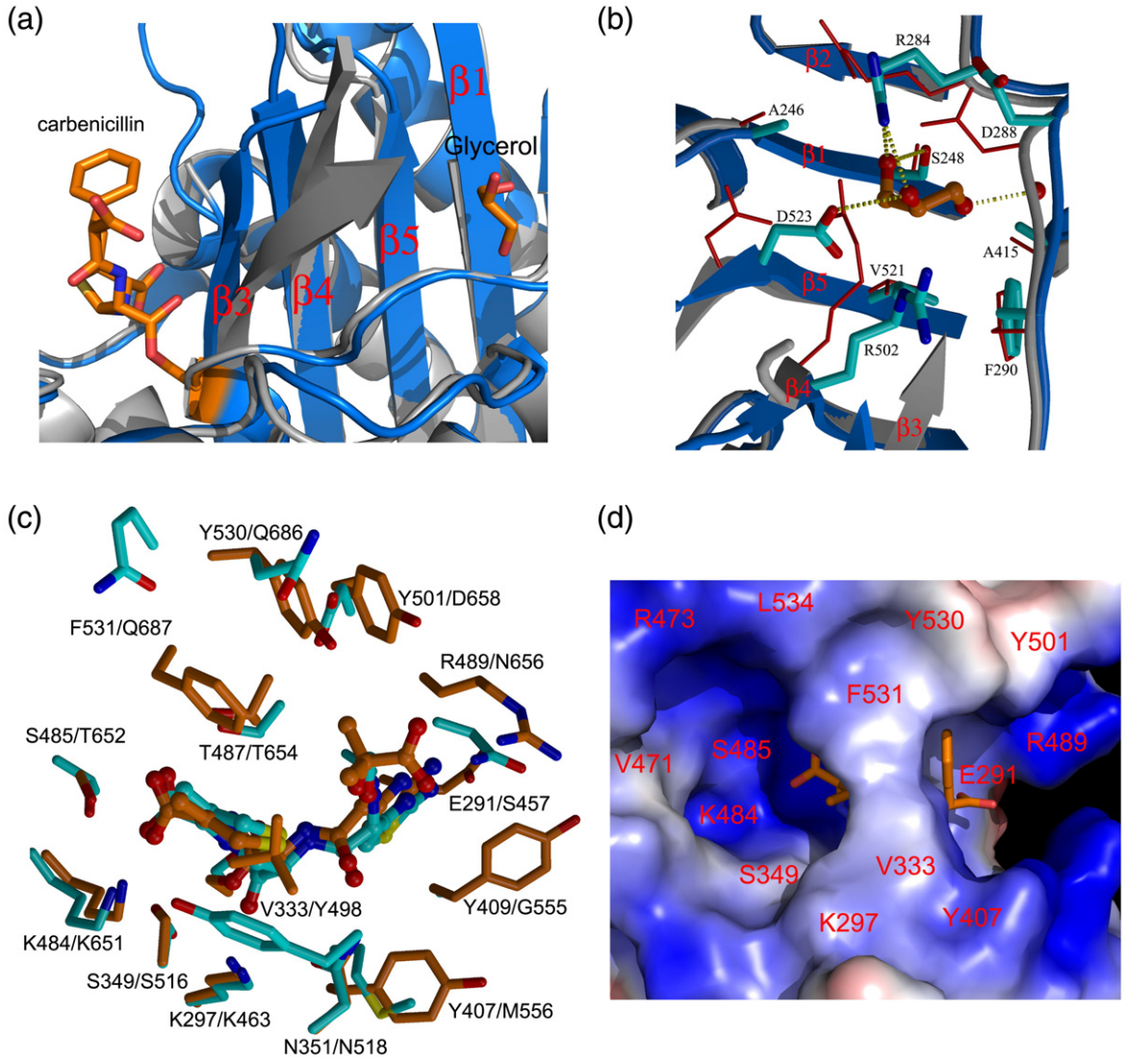


Fig. 4. Glycerol binding, comparison with PBP1b, and electrostatic properties of the active site. (a) A diagram showing that the glycerol binding site is separated from the inhibitor binding site by β3–β4 strands and large conformational changes of the two strands due to inhibitor binding. The main-chain backbones of apo-PBP3 and the PBP3/carbenicillin complex are shown as gray and blue ribbons, respectively. (b) Comparison of the glycerol-binding site in the PBP3/carbenicillin complex with that in the apo-enzyme. The main chains and side chains are shown as ribbons and sticks, respectively (in gray and red for apo-PBP3; in blue and cyan for the carbenicillin complex), and the glycerol molecule is drawn as orange sticks. (c) Comparison of the active sites of the PBP3/ceftazidime complex (orange) and the *S. pneumoniae* PBP1b/cefotaxime complex (cyan). (d) Electrostatic surface showing the inhibitor-binding pocket of the PBP3/carbenicillin complex. (a) and (c) were produced with PyMOL (www.pymol.org).

Active site flexibility and variations

A common feature observed in the crystal structures of PBPs is the flexibility of the active sites, especially the loop linking β3–β4. This may be an intrinsic property of these enzymes that is required for substrate binding and release during the catalytic process. Active site flexibility has also been suggested to be a key mechanism for acquisition of drug resistance.[11] Comparison of the structure of antibiotic-sensitive wild-type PBPs with those from resistant strains has shown that mutations occur in residues close to but not actually in the active site.[4,11] The conformational change in β3 and β4 strands introduced by inhibitor binding opens up a small pocket on the other side of the β-sheet that is occupied by water molecules in apo-PBP3 (Fig. 4a). The crystal of the PBP3/carbenicillin complex had been soaked with cryoprotectant solution containing 20% glycerol prior to freezing (no cryoprotectant was used for apo and ceftazidime-soaked crystals). As a result, the pocket is

occupied by a well-ordered glycerol molecule, which makes hydrogen-bonding interactions with S248, R284, D523, and a trapped water molecule, and hydrophobic contacts with residues G247, S248, F290, and V521 (Fig. 4b). This pocket could be targeted for the design of a new type of inhibitor that restricts active site "breathing." The influence of such a ligand could be to increase the free energy of PBP3 ligand binding by restricting the conformational flexibility of the enzyme.[27] A potential benefit could be use in conjunction with β-lactams to combat drug resistance.

Apart from the three conserved active site motifs, residues lining the substrate-binding cleft of PBPs are poorly conserved. Amino acid insertions and deletions often occur among PBPs and result in variations in the shape and size of the active site clefts. For example, *S. pneumoniae* PBP2x has an extended β3–β4 loop and a shorter linker between β5 and α11, resulting in an active-site cleft that is wider at one end and narrower at the other end compared to PBP3. The structure of the complex of *S. pneumoniae* PBP1b with cefotaxime, a close analogue of ceftazidime, has been reported,[11] enabling a direct comparison of binding modes with the PBP3 complex. Superimposition of the complexes reveals that the two inhibitors dock into their respective targets in a remarkably similar way. The orientations of the residues that form hydrogen bonds with the common β-lactam core of the inhibitors, namely S485 (PBP3)/T652 (PBP1b), T487 (PBP3)/T654 (PBP1b), and N351 (PBP3)/N518 (PBP1b), are closely matched (Fig. 4c). N656 (PBP1b) and R489 (PBP3), which interact with the R_1 groups of cefotaxime and ceftazidime, respectively, are also closely aligned, with their relative positions determined by the size of the R_1 substituents. These observations point to a common binding mode for cephalosporin analogues to the HMM PBPs. Residues E291, Y409, and R489 in PBP3, which make key interactions with the R_1 groups of ceftazidime, are replaced in PBP1b by S457, G558, and N656, respectively, likely reducing the affinity of PBP1b for the inhibitor given the different nature of these residues. More generally, sequence variation at these three positions could be a major factor contributing to the different specificities and affinities of cephalosporin analogues across PBPs.

Finally, it is striking that in all published structures of PBP/inhibitor complexes, the inhibitors occupy the central portion of the active site pocket, leaving a large volume of the cleft unoccupied (Fig. 4c and d). It may be possible to design inhibitors that utilize these "free" volumes to increase protein–inhibitor interactions and/or to prevent β-lactamase hydrolysis. Such modifications offer the potential of enhancing binding affinity for the target and of overcoming drug resistance caused by β-lactamase production.

Experimental Procedures

Protein production

A truncated version of the *ftsI* gene, omitting the bases encoding the first 34 amino acids (residues D35-G579), was amplified by PCR using *P. aeruginosa* PAO1 strain genomic DNA (ATCC strain 15692; LGC Standards Office, UK). The template and primers were designed for ligation-independent cloning *via* In-fusion™ technology (Clontech). PCR with KOD HiFi™ polymerase (Merck) used the following forward and reverse primers: 5′-aagttctgtttcagggcccgGACCTGCACGTGATCGACC-3′ and 5′-atggtctagaaagctttaGCCACGCCCTCCTTTTGC-3′, respectively. The PCR product was purified using paramagnetic microparticles[28] (Agencourt AmPure™ System; Beckman Coulter) and inserted into the vector pOPINF using the protocol of Berrow *et al.*[29] Recombinant clones were identified by PCR with the gene-specific forward primer and a T7 reverse primer, and verified by DNA sequencing. Native PBP3 was produced in the *E. coli* strain Rosetta *p*LysS (DE3) using autoinduction.[30] Cells were grown in 2-L cultures of Overnight™ Express Instant TB media (Merck) at 37 °C for 4 h, and the temperature was then lowered to 25 °C. After incubation for a further 20 h, the cells were harvested by centrifugation for 15 min at 6000*g* and lysed using a Basic-Z Cell Disruptor (Constant Systems) at 30,000 psi in the presence of 50 mM Tris–HCl (pH 7.5), 500 mM NaCl, 0.2% Tween-20, 10 μg/ml DNase, and an ethylenediaminetetraacetic-acid-free Protease Inhibitor Cocktail Tablet (Roche). The lysate was centrifuged at 16,000*g* for 30 min to remove cell debris before the soluble fraction was loaded onto a 1-ml HisTrap FF column (GE Healthcare). The column was washed with 50 mM Tris–HCl (pH 7.5), 500 mM NaCl, and 20 mM imidazole before the elution of the protein in 50 mM Tris–HCl (pH 7.5), 500 mM NaCl, and 500 mM imidazole. The protein was then injected onto a 16/60 HiLoad™ Superdex 75 column (GE Healthcare) and eluted in 20 mM Tris–HCl (pH 7.5) and 200 mM NaCl. Protein-containing fractions were analyzed by SDS-PAGE (NuPage; Invitrogen) and combined, and the NaCl concentration was adjusted to 400 mM to improve protein solubility. Protein (4 mg/ml) was flash frozen and stored at −80 °C.

Thermal shift assay

A thermal shift assay,[31,32] using the fluorescent dye SYPRO Orange (Invitrogen), was employed to demonstrate that the recombinant PBP3 binds the β-lactams and to determine their effect on protein stability. PBP3 was screened with carbenicillin and ceftazidime. Solutions of PBP3 at 4 μM, 1000× SYPRO Orange, and compounds at 500 μM concentration were made, and 40 μl was added to a 96-well thin-wall PCR plate (Thermo Scientific). The plates were sealed with adhesive PCR seal (4titude) and heated in an Mx3005p qPCR machine (Stratagene) from 25 °C to 95 °C at a rate of 1 °C/min. Fluorescence changes were monitored with excitation and emission wavelengths of 492 nm and 610 nm, respectively. Reference wells (i.e., solutions consisting only of PBP3 and the dye) were used to compare the melting temperature (T_m)

values. In order to rule out any effects caused by the presence of buffer, we used wells with buffer and 1000× SYPRO Orange as controls. Experiments were carried out in triplicate, and nonlinear regression analysis was determined in the curve-fitting program Prism (GraphPad Software). T_m values were calculated for each well and compared to the reference T_m values to obtain ΔT_m for each compound. A Microsoft Excel script† was used for convenient observations and comparison of melting curves and thermal shifts.

Crystallization, data collection, and structure determination

Protein crystallization was carried out using standard Oxford Protein Production Facility protocols.[33] Native crystals were grown by sitting-drop vapor diffusion at room temperature in 25% (wt/vol) polyethylene glycol 3350 and 0.1 M 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol propane (pH 7.8) containing 1% (wt/vol) protamine sulfate (Silver Bullet Screen; Hampton Research). Acyl complexes were obtained by either cocrystallization or soaking experiments. Crystals of the carbenicillin complex were obtained in 20% (wt/vol) polyethylene glycol 3350 and 0.2 M disodium tartrate containing 0.5 mM carbenicillin. The PBP3/ceftazidime complex was obtained by soaking native crystals in the crystallization solution containing 10 mM ceftazidime for 30 min at room temperature. All crystals were flash frozen in liquid nitrogen and then kept at −173 °C under a stream of nitrogen gas during data collection. No cryoprotectant was used apart from the carbenicillin complex, where a crystal was soaked in solution containing 20% glycerol (vol/vol) and 80% of the reservoir solution for a few seconds before freezing. X-ray data were collected for the carbenicillin complex on beamline ID23.2 of the European Synchrotron Radiation Facility (ESRF; Grenoble, France), and data sets from the native and ceftazidime-soaked crystals were collected on beamlines I03 and I02 of the Diamond Light Source (Harwell, UK), respectively.

The diffraction data were indexed and integrated with DENZO, and merged with SCALEPACK.[34] The statistics of the data are shown in Table 1. The crystals belong to the orthorhombic space group $P2_12_12_1$, with one protein molecule per asymmetric unit and an approximate bulk solvent content of 44%. The structure of the PBP3/carbenicillin complex was solved by molecular replacement using the coordinates of *N. gonorrhoeae* PBP2 (PDB code 3equ)[4] as search model. The starting model shares a 34% sequence identity with *P. aeruginosa* PBP3 and lacks the 72 residues that form the head subdomain. The orientation and position of the model in the PBP3 crystal were determined using the program MOLREP[35] and refined with rigid-body refinement using data from 30 Å to 4.0 Å. The model was then subjected to restrained maximum-likelihood refinement with REFMAC using all data up to 2.3 Å resolution.[36] The $2F_o - F_c$ map calculated at this stage showed well-defined electron density for the transpeptidase domain, which allowed the majority of residues within the domain to be built with confidence; the electron density for the N-terminal domain, especially for the missing head subdomain, was poor and fragmented, but gradually improved after each round of refinement and model rebuilding with Coot.[37] Several rounds of refinement and model rebuilding led to a model of 498 residues (out of 543), with good stereochemistry and an *R*-factor of 0.196 (R_{free} = 0.250) for all data up to 2.3 Å resolution. The structure of the PBP3/carbenicillin complex was then used to determine the structures of the apoenzyme and the ceftazidime complex. The statistics for the final refined structures are given in Table 1.

† ftp://ftp.sgc.ox.ac.uk/pub/biophysics

Accession numbers

The atomic coordinates and structure factors for apo-PBP3, the PBP3/carbenicillin complex, and the PBP3/ceftazidime complex have been deposited in the PDB under accession codes 3oc2, 3ocl, and 3ocn, respectively.

Supplementary materials related to this article can be found online at doi:10.1016/j.jmb.2010.10.024

Acknowledgements

The Oxford Protein Production Facility was funded by the Medical Research Council and the Biotechnology and Biological Research Council. D.I. S. was supported by the Medical Research Council, and the Wellcome Trust Center for Human Genetics was supported by the Wellcome Trust (grant 075491/Z/04).

Research in Dundee was supported by the Wellcome Trust (grants 082596 and 083481) and the European Commission Seventh Framework Program (FP7/2007-2013; AEROPATH grant 223461). We thank the staff of beamlines I02 and I03 at Diamond Light Source and of beamline ID23-2 at ESRF for help with data collection.

References

1. Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A. & Charlier, P. (2008). The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol. Rev.* **32**, 234–258.
2. Zapun, A., Contreras-Martel, C. & Vernet, T. (2008). Penicillin-binding proteins and beta-lactam resistance. *FEMS Microbiol. Rev.* **32**, 361–385.
3. Sung, M. T., Lai, Y. T., Huang, C. Y., Chou, L. Y., Shih, H. W., Cheng, W. C. *et al.* (2009). Crystal structure of the membrane-bound bifunctional transglycosylase PBP1b from *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **106**, 8824–8829.
4. Powell, A. J., Tomberg, J., Deacon, A. M., Nicholas, R. A. & Davies, C. (2009). Crystal structures of penicillin-binding protein 2 from penicillin-susceptible and -resistant strains of *Neisseria gonorrhoeae* reveal an unexpectedly subtle mechanism for antibiotic resistance. *J. Biol. Chem.* **284**, 1202–1212.

5. Macheboeuf, P., Di Guilmi, A. M., Job, V., Vernet, T., Dideberg, O. & Dessen, A. (2005). Active site restructuring regulates ligand recognition in class A penicillin-binding proteins. *Proc. Natl Acad. Sci. USA*, **102**, 577–582.
6. Yamada, M., Watanabe, T., Baba, N., Takeuchi, Y., Ohsawa, F. & Gomi, S. (2008). Crystal structures of biapenem and tebipenem complexed with penicillin-binding proteins 2X and 1A from *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **52**, 2053–2060.
7. Gordon, E., Mouz, N., Duee, E. & Dideberg, O. (2000). The crystal structure of the penicillin-binding protein 2x from *Streptococcus pneumoniae* and its acyl-enzyme form: implication in drug resistance. *J. Mol. Biol.* **299**, 477–485.
8. Dessen, A., Mouz, N., Gordon, E., Hopkins, J. & Dideberg, O. (2001). Crystal structure of PBP2x from a highly penicillin-resistant *Streptococcus pneumoniae* clinical isolate: a mosaic framework containing 83 mutations. *J. Biol. Chem.* **276**, 45106–45112.
9. Pares, S., Mouz, N., Petillot, Y., Hakenbeck, R. & Dideberg, O. (1996). X-ray structure of *Streptococcus pneumoniae* PBP2x, a primary penicillin target enzyme. *Nat. Struct. Biol.* **3**, 284–289.
10. Pernot, L., Chesnel, L., Le Gouellec, A., Croize, J., Vernet, T., Dideberg, O. & Dessen, A. (2004). A PBP2x from a clinical isolate of *Streptococcus pneumoniae* exhibits an alternative mechanism for reduction of susceptibility to beta-lactam antibiotics. *J. Biol. Chem.* **279**, 16463–16470.
11. Contreras-Martel, C., Dahout-Gonzalez, C., Martins Ados, S., Kotnik, M. & Dessen, A. (2009). PBP active site flexibility as the key mechanism for beta-lactam resistance in pneumococci. *J. Mol. Biol.* **387**, 899–909.
12. Lim, D. & Strynadka, N. C. (2002). Structural basis for the beta lactam resistance of PBP2a from methicillin-resistant *Staphylococcus aureus*. *Nat. Struct. Biol.* **9**, 870–876.
13. Lyczak, J. B., Cannon, C. L. & Pier, G. B. (2002). Lung infections associated with cystic fibrosis. *Clin. Microbiol. Rev.* **15**, 194–222.
14. Stover, K. C., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J. *et al.* (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1: an opportunistic pathogen. *Nature*, **406**, 959–964.
15. Gotoh, N., Nunomura, K. & Nishino, T. (1990). Resistance of *Pseudomonas aeruginosa* to cefsulodin: modification of penicillin-binding protein 3 and mapping of its chromosomal gene. *J. Antimicrob. Chemother.* **25**, 513–523.
16. O'Callaghan, C. H., Acred, P., Harper, P. B., Ryan, D. M., Kirby, S. M. & Harding, S. M. (1980). GR20263, a new broad-spectrum cephalosporin with anti-pseudomonal activity. *Antimicrob. Agents Chemother.* **17**, 876–883.
17. Godfrey, A. J., Bryan, L. E. & Rabin, H. R. (1981). Beta-lactam-resistant *Pseudomonas aeruginosa* with modified penicillin-binding proteins emerging during cystic fibrosis treatment. *Antimicrob. Agents Chemother.* **19**, 705–711.
18. Davies, T. A., Shang, W., Bush, K. & Flamm, R. K. (2008). Affinity of doripenem and comparators to penicillin-binding proteins in *Escherichia coli* and *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **52**, 1510–1512.
19. Georgopapadakou, N. H. (1993). Penicillin-binding proteins and bacterial resistance to Δ-lactams. *Antimicrob. Agents Chemother.* **37**, 2045–2053.
20. Liao, X. & Hancock, R. E. (1997). Susceptibility to beta-lactam antibiotics of *Pseudomonas aeruginosa* over-producing penicillin-binding protein 3. *Antimicrob. Agents Chemother.* **41**, 1158–1161.
21. Acred, P., Brown, D. M., Knudsen, E. T., Rolinson, G. N. & Sutherland, R. (1967). New semi-synthetic penicillin active against *Pseudomonas pyocyanea*. *Nature*, **215**, 25–30.
22. Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics*, **15**, 132–134.
23. Stuart, D. I., Levine, M., Muirhead, H. & Stammers, D. K. (1979). Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 Å. *J. Mol. Biol*, **134**, 109–142.
24. Macheboeuf, P., Contreras-Martel, C., Job, V., Dideberg, O. & Dessen, A. (2006). Penicillin binding proteins: key players in bacterial cell cycle and drug resistance processes. *FEMS Microbiol. Rev.* **30**, 673–691.
25. Chen, Y., Zhang, W., Shi, Q., Hesek, D., Lee, M., Mobashery, S. & Shoichet, B. K. (2009). Crystal structures of pencillin-binding protein 6 from *Escherichia coli*. *J. Am. Chem. Soc.* **131**, 14345–14354.
26. Zhou, M. & Notari, R. E. (1995). Influence of pH, temperature, and buffers on the kinetics of ceftazidime degradation in aqueous solutions. *J. Pharm. Sci.* **84**, 534–538.
27. Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. (2007). Conformational entropy in molecular recognition by proteins. *Nature*, **448**, 325–329.
28. DeAngelis, M. M., Wang, D. G. & Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.* **23**, 4742–4743.
29. Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N. *et al.* (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45.
30. Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif.* **41**, 207–234.
31. Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E. *et al.* (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screening*, **6**, 429–440.
32. Lo, M. C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M. & Ellestad, G. (2004). Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **332**, 153–159.
33. Walter, T. S., Diprose, J. M., Mayo, C. J., Siebold, C., Pickford, M. G., Carter, L. *et al.* (2005). A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr. Sect. D*, **61**, 651–657.
34. Otwinowski, Z. M. W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.

35. Vagin, A. A. T. A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025.
36. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D*, **53**, 240–255.
37. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.

APPENDIX B

Rao VA, Shepherd SM, English G, Coulthurst SJ and Hunter WN (2011) The structure of *Serratia marcescens* Lip, a membrane-bound component of the type VI secretion system. ***Acta Crystallographica*** **D67: 1065-1072**

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure of *Serratia marcescens* Lip, a membrane-bound component of the type VI secretion system

Vincenzo A. Rao, Sharon M. Shepherd, Grant English, Sarah J. Coulthurst and William N. Hunter*

College of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland

Correspondence e-mail: w.n.hunter@dundee.ac.uk

Received 21 September 2011
Accepted 2 November 2011

PDB Reference: Lip, 4a1r.

Lip is a membrane-bound lipoprotein and a core component of the type VI secretion system found in Gram-negative bacteria. The structure of a Lip construct (residues 29–176) from *Serratia marcescens* (*Sm*Lip) has been determined at 1.92 Å resolution. Experimental phases were derived using a single-wavelength anomalous dispersion approach on a sample cocrystallized with iodide. The membrane localization of the native protein was confirmed. The structure is that of the globular domain lacking only the lipoprotein signal peptide and the lipidated N-terminus of the mature protein. The protein fold is dominated by an eight-stranded β-sandwich and identifies *Sm*Lip as a new member of the transthyretin family of proteins. Transthyretin and the only other member of the family fold, 5-hydroxyisourate hydrolase, form homo-tetramers important for their function. The asymmetric unit of *Sm*Lip is a tetramer with 222 symmetry, but the assembly is distinct from that previously noted for the transthyretin protein family. However, structural comparisons and bacterial two-hybrid data suggest that the *Sm*Lip tetramer is not relevant to its role as a core component of the type VI secretion system, but rather reflects a propensity for *Sm*Lip to participate in protein–protein interactions. A relatively low level of sequence conservation amongst Lip homologues is noted and is restricted to parts of the structure that might be involved in interactions with physiological partners.

1. Introduction

Protein secretion systems are critical to the virulence and host-interaction processes of Gram-negative pathogens. Different bacterial species possess different combinations of one or more specialized proteinaceous machines that secrete toxins, adhesins, hydrolytic enzymes and proteins able to manipulate eukaryotic signalling pathways (Gerlach & Hensel, 2007; Holland, 2010). The most recently discovered system, the type VI secretion system (T6SS), is present in many Gram-negative bacteria and is implicated in virulence in important human pathogens including *Pseudomonas aeruginosa* (Cascales, 2008; Filloux *et al.*, 2008; Jani & Cotter, 2010). It has also been shown to contribute to the virulence of economically significant animal and plant pathogens (Liu *et al.*, 2008; Blondel *et al.*, 2010; Sarris *et al.*, 2010). Some T6SSs appear to target other bacterial cells instead of, or in addition to, eukaryotic cells (Hood *et al.*, 2010; MacIntyre *et al.*, 2010; Murdoch *et al.*, 2011). This suggests that T6SSs may contribute to allowing pathogens to proliferate in polymicrobial infection sites and/or to persist in different environmental reservoirs (Schwarz *et al.*, 2010; Murdoch *et al.*, 2011). *Serratia marcescens* is an opportunistic

pathogen, a significant cause of hospital-acquired infections and an important reservoir of antibiotic-resistance determinants in the clinical environment (Hejazi & Falkiner, 1997). It is also a tractable model organism in which to dissect the structure–function relationships in the T6SS (Murdoch *et al.*, 2011).

Studies of the T6SS have started to reveal information on the components and the biological role of this recently discovered system (Cascales, 2008; Filloux *et al.*, 2008; Pukatzki *et al.*, 2009; Bönemann *et al.*, 2010). T6SSs are large multiprotein complexes encoded on variable gene clusters characterized by the presence of genes encoding 13 'core' components. These are thought to form the basic secretion apparatus, which is coupled with 'accessory' components that are conserved across many or only a few systems. Key core components include the putative extracellular Hcp/VgrG assembly, which is thought to form a cell-puncturing device similar to that of bacteriophage tail structures (Pukatzki *et al.*, 2009). There are a number of predicted cytoplasmic proteins (*e.g.* an ATPase called ClpV) and several inner membrane proteins (*e.g.* IcmF and IcmH). Additionally, and the subject of this work, the only outer membrane component reported to date is a periplasmic-facing outer membrane lipoprotein (Lip; Aschtgen *et al.*, 2008).

Genetic studies indicate that in *S. marcescens* this lipoprotein (*Sm*Lip) makes an essential contribution to the basic function of the T6SS and to T6SS-dependent antibacterial killing activity (Murdoch *et al.*, 2011). We now report the high-resolution structure of *Sm*Lip determined following phase determination using single-wavelength anomalous dispersion (SAD) measurements based on the scattering properties of iodide ions. The localization of the protein in *S. marcescens* itself and bacterial two-hybrid data are reported to investigate the propensity for self-association. The structure reveals a remarkable similarity to transthyretin, a vertebrate hormone-distribution protein, and comparisons suggest which parts of *Sm*Lip may be involved in protein–protein interactions with partner components of the T6SS.

2. Methods

2.1. Protein expression and purification

The *S. marcescens lip* gene (*SMA2252*; Murdoch *et al.*, 2011) encoding amino-acid residues 30–176 was amplified from genomic DNA (strain Db10) using the forward primer 5′-catatgGCCAAAAGCGTGCCGTCGCGTTACAG-3′ and the reverse primer 5′-ggatccTCAGTCGACCTTTTTTACG-GGGCGCAGGC-3′ (the lower-case sequences correspond to the *Nde*I/*Bam*HI restriction sites used for cloning). The PCR product was ligated into PCR-BluntII-TOPO using the Zero Blunt TOPO Cloning Kit (Invitrogen) and then cloned into a pET15b (Novagen) cloning vector modified to encode a tobacco etch virus (TEV) protease cleavage site in place of the thrombin protease cleavage site. The construct was verified by DNA sequencing (DNA Sequencing Unit, University of Dundee).

The recombinant protein was produced in *Escherichia coli* BL21 (DE3) pLysS cells (Stratagene). Cultures were grown for 3 h at 310 K in auto-induction medium (Studier, 2005) supplemented with 50 $\mu g\,l^{-1}$ carbenicillin before overnight growth at 295 K. Cells were harvested by centrifugation (3500*g* at 277 K for 30 min). The cell pellet was resuspended in buffer *A* (25 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl, 20 m*M* imidazole pH 8.5) supplemented with an EDTA-free Protease Inhibitor Cocktail Tablet (Roche) and 0.2 mg DNase I (Sigma–Aldrich). Cells were lysed using a continuous-flow cell disrupter (Constant Systems) at 207 MPa and cell debris was removed following centrifugation (40 000*g* at 277 K for 30 min). *Sm*Lip was purified using nickel-affinity chromatography with a 5 ml HisTrap HP column (GE Healthcare) precharged with Ni^{2+}. A step gradient of 5% buffer *B* (25 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl, 500 m*M* imidazole) was used to remove histidine-rich proteins. A linear concentration gradient of imidazole from 5 to 50% buffer *B* was applied to elute the product, which was then dialyzed against buffer *C* (25 m*M* Tris–HCl, 250 m*M* NaCl pH 7.5) at 277 K overnight in the presence of His-tagged TEV protease. The resulting mixture was applied onto the HisTrap column, which bound the cleaved His tag, TEV protease and uncleaved *Sm*Lip. The *Sm*Lip sample from which the His tag had been cleaved was present in the flowthrough. Fractions were analyzed using SDS–PAGE and those containing *Sm*Lip were pooled. The protein was further purified by size-exclusion chromatography using a Superdex 75 26/60 column (GE Healthcare) equilibrated with buffer *C* on an ÄKTApurifier (GE Healthcare). The column had previously been calibrated with the molecular-weight standards blue dextran (>2000 kDa), thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29.5 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa) (GE Healthcare; data not shown). The protein eluted as one peak of approximate mass 17 kDa, corresponding to a monomer. Fractions containing the protein were pooled and concentrated to 10 $mg\,ml^{-1}$ using Amicon Ultra devices (Millipore) for subsequent use. The purity of the protein was confirmed by SDS–PAGE and mass spectrometry (Fingerprint Proteomics Facility, University of Dundee). A theoretical extinction coefficient of 16 960 $M^{-1}\,cm^{-1}$ at 280 nm was used to estimate the protein concentration (*ProtParam*; Gasteiger *et al.*, 2005); the theoretical mass of one subunit was estimated as 16.1 kDa with a calculated isoelectric point of 5.4. The purified protein sample was stored at 277 K until further use.

2.2. Crystallization, data collection and structure determination

Initial crystallization screens were carried out at 293 K by the sitting-drop vapour-diffusion method in 96-well plates. This was achieved using a Phoenix liquid-handling system (Rigaku, Art Robbins Instruments) and the commercially available PEG (Qiagen) and JCSG+ (Molecular Dimensions) screens. Crystallization occurred in two conditions, which were further optimized using the hanging-drop vapour-diffusion

method with drops consisting of 1 µl protein solution at 10 mg ml^{-1} in 25 m*M* Tris–HCl pH 7.5, 250 m*M* NaCl and 1 µl reservoir solution. The two conditions involved reservoirs consisting of 20% polyethylene glycol 3350, 200 m*M* KI and of 15% polyethylene glycol 3350, 200 m*M* NaCl. Monoclinic blocks with minimum dimensions of approximately 0.3 mm grew over 2 d and the addition of glycerol to 10% proved to be a suitable cryoprotectant.

Crystals from the iodide-containing condition were characterized first and data set I was measured in-house using a Rigaku MicroMax-007 rotating-anode X-ray generator (Cu $K\alpha$, $\lambda = 1.541$ Å) coupled to an R-AXIS IV^{++} image-plate detector. A crystal from the second condition was stored in liquid N_2 and subsequently used to measure a high-resolution data set (data set II) on beamline ID29 at the European Synchrotron Radiation Facility (ESRF; Grenoble, France) using an ADSC Q315R detector. All data were indexed and integrated using *XDS* (Kabsch, 2010) and scaled using *SCALA* (Evans, 2006) from the *CCP*4 program suite (Winn *et al.*, 2011).

Data set I was used to solve the structure by SAD methods targeting the iodides present in the crystallization conditions and to acquire a fairly complete model. The sites of potential anomalous scattering ions or atoms were identified using *PHENIX* (Adams *et al.*, 2010) and experimental phases were calculated using *Phaser* (McCoy *et al.*, 2007). Density modification was carried out using histogram matching, averaging on the basis of noncrystallographic symmetry (NCS), and model building was carried out using *RESOLVE* (Terwilliger, 2003). NCS restraints were employed in the initial refinement calculations, which were performed using *REFMAC*5 (Murshudov *et al.*, 2011). Inspection of the model and the fit to electron-density and difference density maps was carried out in *Coot* (Emsley *et al.*, 2010). The analysis then switched to the high-resolution synchrotron data set II when it became available and this was used to complete the refinement. *MolProbity* (Chen *et al.*, 2010) was used to investigate model geometry in combination with the validation tools provided in *Coot*. Analyses of surface areas and interactions were made using the *PISA* (Krissinel & Henrick, 2007) web service and secondary-structure analysis was performed using *DSSP* (Kabsch & Sander, 1983). Crystallographic statistics are summarized in Table 1.

Table 1
Crystallographic statistics.

Values in parentheses are for the highest resolution shell.

	Data set I	Data set II
Space group	*C*2	*C*2
Wavelength (Å)	1.5418	1.007
Unit-cell parameters (Å, °)	$a = 139.7$, $b = 77.6$, $c = 54.3$, $\beta = 98.4$	$a = 139.7$, $b = 77.8$, $c = 54.5$, $\beta = 98.3$
Resolution (Å)	19.7–2.35 (2.48–2.35)	39.8–1.92 (2.02–1.92)
No. of reflections recorded	94650 (11192)	280802 (38597)
Unique reflections	23495 (3109)	43478 (6188)
Completeness (%)	98.3 (90.0)	98.2 (95.8)
Multiplicity	4.0 (3.6)	6.5 (6.2)
$\langle I/\sigma(I)\rangle$	30.9 (6.6)	20.6 (3.7)
Anomalous completeness (%)	95.7 (84.5)	—
Anomalous multiplicity	2.0 (1.8)	—
Wilson *B* (Å^2)	47.7	32.7
No. of residues/waters	—	541/336
$R_{\rm merge}$† (%)	2.6 (17.1)	4.9 (44.2)
$R_{\rm work}$‡ (%)	—	22.0
$R_{\rm free}$§ (%)	—	29.2
Average *B* factors (Å^2)		
Chain *A*	—	36.1
Chain *B*	—	41.6
Chain *C*	—	49.5
Chain *D*	—	59.8
Waters	—	46.2
Na^+	—	37.8
Ethylene glycol	—	60.2
Cruickshank DPI¶ (Å)	—	0.2
Ramachandran plot		
Most favoured	—	516 residues
Additional allowed	—	21 residues
Outliers	—	Molecule *D*: Phe97, Asp129; molecule *B*: Pro142, Ser154
R.m.s.d. on ideal values††		
Bond lengths (Å)	—	0.01
Bond angles (°)	—	1.42

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. ‡ $R_{\rm work} = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$, where $F_{\rm obs}$ is the observed structure factor and $F_{\rm calc}$ is the calculated structure factor. § $R_{\rm free}$ is the same as $R_{\rm cryst}$ except calculated with a subset (5%) of data that were excluded from the refinement calculations. ¶ Cruickshank (1999). †† Engh & Huber (1991).

2.3. Bacterial two-hybrid analyses

For generation of the plasmid pSC072, the gene fragment encoding *Sm*Lip amino acids 27–176 was PCR-amplified using primers 5′-TATAgcatgcGTAAAGAGGAGGCTGCATGTC-TTCCGCCAAAAGC-3′ and 5′-TATAtctagaGAGTCGAC-CTTTTTTACGGGGC-3′ and cloned into the vector pUT18 (Karimova *et al.*, 2001) using *Sph*I and *Xba*I restriction sites. The restriction sites are shown in lower case. For generation of another plasmid, pSC080, the same gene fragment was PCR-amplified using primers 5′-TATAggatccAATGTCTTC-CGCCAAAAGCG-3′ and 5′-TATAggtaccAATGATGACG-ACCCCTATCGC-3′ and cloned into vector pT25 (Karimova *et al.*, 1998) using *Bam*HI and *Kpn*I restriction sites (again shown in lower case). Bacterial two-hybrid analyses were performed following established protocols (Karimova *et al.*, 1998, 2000). *E. coli* BTH101 was transformed with pSC072 (or pUT18 control) and pSC080 (or pT25 control) and the colour of the resulting transformants was scored on MacConkey media with 0.2% maltose (with a positive result being red). For quantitative measurement of the interaction, β-galactosidase assays were performed as described by Murdoch *et al.* (2011) on double-transformed BTH101 grown at 303 K in Luria–Bertani broth (LB) and permeabilized with toluene. Replicate assays were performed on independent transformants.

2.4. Cellular localization of Lip

Wild-type *S. marcescens* strain Db10 and the *lip* mutant SJC10 (Murdoch *et al.*, 2011) were grown for 8 h at 303 K

in LB. Subcellular fractionation was performed following an established method (Hatzixanthis *et al.*, 2003). In brief, following isolation of clean supernatant by centrifugation, washed cell pellets were resuspended in 50 m*M* Tris–HCl pH 7.5, 40%(*w*/*v*) sucrose at 10 ml per gram of cells. EDTA was then added to 5 m*M* (final concentration) and lysozyme was added to 0.6 mg ml^{-1} before incubation at 310 K for 30 min. Sphaeroplasts were harvested by centrifugation and taken up in an equivalent volume of 50 m*M* Tris–HCl pH 7.5 before French pressure treatment. Following ultracentrifugation of the resultant crude extract, the isolated membranes were again taken up in an equivalent volume of 50 m*M* Tris–HCl pH 7.5. This protocol ensured that equivalent proportions of each cell fraction were assayed. 4 µl of each fraction was mixed with SDS sample buffer (100 m*M* Tris–HCl pH 6.8, 3.2% SDS, 3.2 m*M* EDTA, 16% glycerol, 0.2 mg ml^{-1} Bromophenol blue, 2.5% β-mercaptoethanol) and separated by 15% SDS–PAGE prior to anti-Lip immunoblotting. Whole-cell samples comparing wild-type *versus* SJC10 were prepared by resuspending cells from 100 ml culture in 100 µl SDS sample buffer and boiling for 5 min prior to loading 6 µl onto the gel. Following SDS–PAGE, proteins were electroblotted onto polyvinylidine fluoride membrane (Millipore). *Sm*Lip was detected by hybridization of the primary antibody polyclonal rabbit anti-Lip (1:4000) followed by the secondary antibody HRP-conjugated goat anti-rabbit (Thermo; 1:10000) and the use of an enhanced chemiluminescent detection kit (Millipore).

3. Results and discussion

3.1. Structure determination

Full-length *Sm*Lip consists of 176 residues. A truncated version of *Sm*Lip consisting of an N-terminal hexahistidine tag plus a TEV protease recognition site followed by residues Ala30–Asp176 was obtained in recombinant form and purified. The N-terminal 29 amino acids, which include the lipidation signal peptide and the first four residues of the mature protein, have been omitted. This sample gave monoclinic crystals. The asymmetric unit consists of four polypeptide chains, labelled *A*–*D*, with an estimated solvent content of 45% and a $V_{\rm M}$ of 2.27 Å^3 Da^{-1}.

Medium-resolution diffraction data were recorded in-house and the anomalous scattering information was used in a SAD approach to phasing. 13 potential iodide positions were identified and produced a figure of merit of 0.43 to 2.35 Å resolution. Subsequently, 12 of these positions were confirmed by refinement with this data set. The initial model constructed in *RESOLVE* consisted of 293 residues, with a correlation coefficient of 0.55 and $R_{\rm work}$ and $R_{\rm free}$ values of 46% and 49%, respectively. The first round of model building in *Coot* extended this to 467 residues, with a correlation coefficient of 0.72 and $R_{\rm work}$ and $R_{\rm free}$ values of 33% and 37%, respectively. At this point the high-resolution synchrotron data (1.92 Å resolution) became available and were used to continue the analysis. The refinement proceeded with the release of NCS restraints and the incorporation of water molecules, an Na^+ ion, ethylene glycol and a number of side chains with dual rotamer conformations. This data set was derived from crystals grown in the presence of chloride instead of iodide. However, we did not assign any chloride ions to the structure, noting that typical water molecules occupy the previously identified iodide-binding sites. The refinement was terminated when there were no significant changes in $R_{\rm work}$ and $R_{\rm free}$ and inspection of the difference density map suggested that no further corrections or additions were justified. Several dual rotamers are incorporated into the model. Disorder was evident at several positions, for example the N-terminus, where it was not possible to interpret diffuse and weak electron density. Consequently, several residues are absent from the model. Molecule *A* consists of residues 32–173; molecule *B* of residues 33–142 and 147–176; molecule *C* of residues 34–50, 53–143 and 147–175; and molecule *D* of residues 33–50 and 55–175. The geometry of the model is acceptable (Table 1).

3.2. Self-association and localization *in vivo*

Previous work on SciN, the Lip homologue from enteroaggregative *E. coli*, showed that the protein is localized in the outer membrane, facing the periplasm (Aschtgen *et al.*, 2008). Examination of the amino-acid sequence of the N-terminus of *Sm*Lip predicts that this is also an outer-membrane lipoprotein. The *LipoP* 1.0 algorithm (Juncker *et al.*, 2003) predicts that *Sm*Lip has a lipoprotein signal peptide and that signal peptidase II cleavage occurs between Gly25 and Cys26, with the cysteine subsequently being lipidated. Additionally, the residue at the +2 position following cleavage is Met27 (*i.e.* it is not an aspartate, which directs retention in the inner

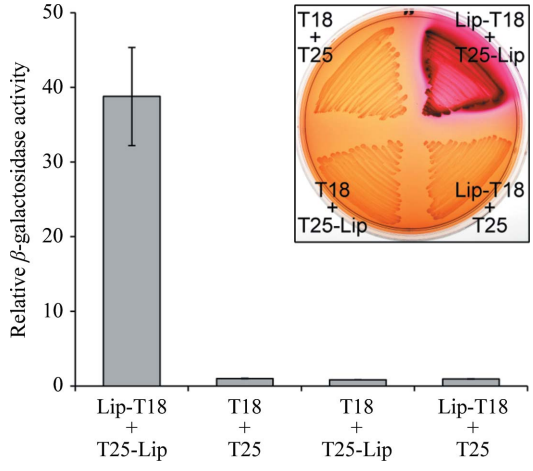


Figure 1
Detection of Lip–Lip self-interaction. The bacterial two-hybrid system was used to detect an *in vivo* interaction between Lip (minus signal peptide) fused to T25 (pSC080) and Lip (minus signal peptide) fused to T18 (pSC072). The empty vectors pUT18 and pT25 represent negative controls. The graph shows the output from the two-hybrid system detected as β-galactosidase activity expressed relative to the pUT18/pT25 baseline level (the maximal β-galactosidase activity observed for the Lip–Lip interaction corresponded to >5000 Miller units). Bars show mean ± SEM. Inset: colourimetric readout of the two-hybrid assay following growth of *E. coli* BTH101 carrying the above plasmids on MacConkey–maltose agar (red is a positive result).

membrane); therefore, *Sm*Lip should proceed to the outer membrane *via* the Lol system (Bos *et al.*, 2007).

In order to investigate whether *Sm*Lip undergoes self-interaction, the bacterial two-hybrid system (Karimova *et al.*, 2000) was utilized in *E. coli*. This assay involves reconstitution of adenylate cyclase activity from two non-interacting cyclase fragments, called T18 and T25, from *Bordetella pertussis*. The presence of cyclic AMP activates the transcription of maltose and lactose catabolic operons by *E. coli*. This can be detected by direct measurement of β-galactosidase activity or by using the observation that bacteria capable of fermenting maltose acidify the medium and thus generate a red colour on MacConkey–maltose indicator plates.

*Sm*Lip was introduced as both bait and prey by encoding on plasmids pUT18 and pT25, and a strong positive result was observed (Fig. 1). Mature *Sm*Lip (lacking the N-terminal signal peptide) was used for this experiment, firstly to correspond to the form of *Sm*Lip for which the structure was solved and secondly to ensure that both partners were localized together in the cytoplasm after fusion with T18 or T25. This positive result indicates that Lip does indeed self-associate within the cell and that neither localization in the outer membrane nor other components of the type VI secretion apparatus are required for self-interaction. We note, however, that this system is unable to distinguish between dimerization or higher order oligomerization.

As a control for any propensity of *Sm*Lip to form non-specific interactions, in addition to the lack of interaction with the T18 and T25 proteins demonstrated in Fig. 1 we tested whether *Sm*Lip gave a positive bacterial two-hybrid result with several cytoplasmic components of the T6SS (with which, as a periplasmic protein, it should not interact). *Sm*Lip gave a negative result (indistinguishable from the T25 + T18 negative control) when tested against the proteins VipB, TssK and TssL (data not shown).

In order to confirm the localization of the native Lip protein in *S. marcescens*, we utilized an anti-Lip polyclonal antibody to probe each of the major cellular fractions in this organism. As shown in Fig. 2, native *Sm*Lip is found exclusively in the membrane fraction, confirming the predicted localization of the protein and the functionality of the signal peptide.

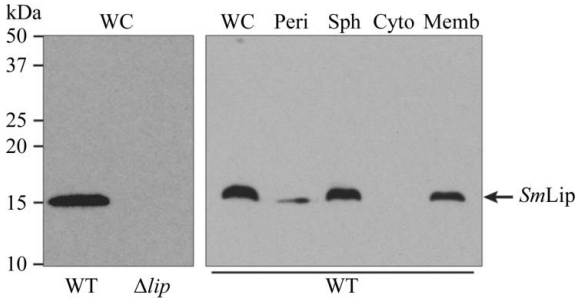


Figure 2
Cellular localization of native *Sm*Lip in *S. marcescens*. Anti-Lip immunoblot of whole cells or cellular fractions prepared from wild-type *S. marcescens* strain Db10 or the Δ*lip* mutant SJC10 (WC, whole cell; Peri, periplasm; Sph, sphaeroplast; Cyto, cytoplasm; Memb, membranes). The predicted size of mature *Sm*Lip is 16 kDa.

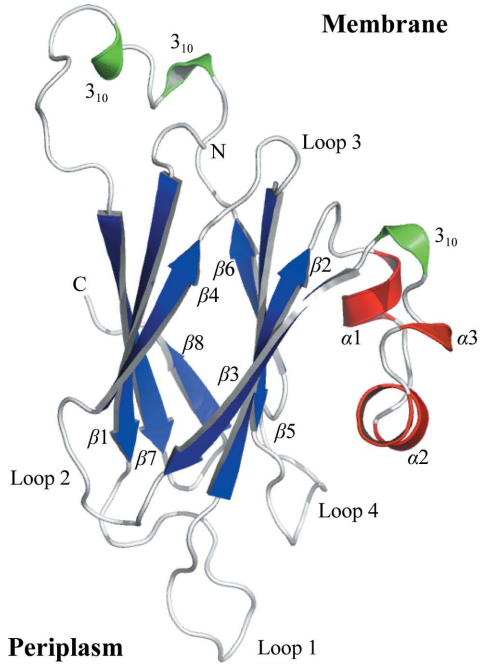


Figure 3
The secondary structure and fold of *Sm*Lip. β-Strands are shown as blue arrows and α-helices and 3_{10}-turns as red and green ribbons, respectively. The N- and C-terminal residues are labelled and the orientation of the protein with respect to the outer membrane and periplasm is suggested.

3.3. Overall structure

The *Sm*Lip polypeptide can be classified as a new member of the transthyretin-like superfamily and a detailed comparison will be given below. The protein displays a compact globular structure dominated by an eight-stranded β-sandwich (Fig. 3; Supplementary Fig. S1[1]). The order of the strands is 8–7–1–4 and 6–5–2–3. There are three short α-helical segments and three 3_{10}-helix turns. The four *Sm*Lip polypeptide chains in the asymmetric unit are similar, with the root-mean-square (r.m.s.) deviations between superimposed C^{α} atoms ranging from 1.3 Å (monomers *A* and *B*) to 0.8 Å (monomers *A* and *D*) with an average value of 0.95 Å.

Although a set of core conserved proteins are encoded by the T6SS gene clusters in different Gram-negative bacteria (data not shown), there is a large degree of variation in the amino-acid sequences of these proteins. Lip and its orthologues, for example, are relatively poorly conserved. Excluding the signal peptide and lipobox motif (Fig. 4), *Sm*Lip shares only about 20% sequence identity with SciN, the homologue from enteroaggregative *E. coli*. This increases to near 40% in comparison with the homologue from the *P. aeruginosa* HSI-1 T6SS. Sequence conservation is noted in loop 1, near α1 and α2, in loop 2 and in the loop 4–β6 region (Fig. 4).

An alignment of *Sm*Lip with eight orthologues (Supplementary Fig. S2[1]) reinforces the observation of a low level of sequence identity for this protein. Excluding two residues

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: MN5005). Services for accessing this material are described at the back of the journal.

in the lipobox motif, only six residues are strictly conserved: Asn48, Leu99-*X*-Pro101-Gly102, Gly120 and Ala124. All six residues appear to contribute to the conformation of specific parts of the fold (data not shown). The side chain of Asn48 accepts a hydrogen bond from the main-chain amide of Gln126, helping to define the conformation of loop 4. The Leu99-*X*-Pro101-Gly102 segment defines the structure of the turn after β3 leading into loop 2. Gly120 and Ala124 occur in β5 and contribute hydrogen bonds to form interactions with β2 and β6 on either side. An increase in size of the side chain at either of these positions would be likely to be disruptive to the formation of this β-sheet, which forms one side of the structure. There is no obvious hydrophobic, basic or acidic surface feature on *Sm*Lip that is likely to be conserved within the Lip proteins since the few conserved residues are mainly buried.

The information provided in §3.2 identifies that the N-terminus of the structure is placed close to the outer membrane, hence the assignment of the orientation of *Sm*Lip with respect to the outer membrane (Fig. 3). By extension, we note that the areas of *Sm*Lip in which sequence conservation is observed mainly appear to contribute to stabilizing parts of the structure that jut out into the periplasm. They may therefore serve to define the structure of parts of Lip that are responsible for interaction with other molecules in the periplasm.

3.4. The tetramer is likely to be a crystallographic artefact

Gel-filtration data acquired during purification indicated that *Sm*Lip is a monomer in solution (data not shown). In contrast, the bacterial two-hybrid data revealed a propensity for self-interaction and the asymmetric unit is a tetramer displaying 222 point-group symmetry (Fig. 5). The accessible surface area (ASA) of the *Sm*Lip polypeptide averages out at approximately 8350 Å^2; the range is from 8200 Å^2 for molecule *D* to 8510 Å^2 for molecule *A*. Each molecule in the asymmetric unit interacts with two of the other three and two types of protein–protein interface are formed between molecules *A*–*B* and *C*–*D* (interface I) and between molecules *A*–*C* and *B*–*D* (interface II). The type I interface, which is larger, covers an area that is approximately 13% of the ASA of the *Sm*Lip molecule. Such coverage certainly indicates potential for a biologically relevant dimer. This interface is primarily formed by the antiparallel alignment of two β7 strands. Three aromatic residues, Phe147, Trp151 and Phe153, contribute van der Waals interactions to the association and, by virtue of their relative bulk, also to the ASA (data not shown). The type II interface covers about 6.5% of the ASA of a molecule, a level typical of the values observed simply owing to molecular packing in a crystal lattice. This interface is formed by the antiparallel alignment of two β4 strands. The areas of *Sm*Lip involved in forming a tetramer are not conserved in the homologues from *E. coli* or *P. aeruginosa* (Fig. 4) and it is unlikely that such a tetramer is a generic feature of this lipoprotein.

The spatial placement of the N-terminal residues in the asymmetric unit is such that it is unlikely that an oligomeric assembly could form when the protein is anchored in the membrane by the lipidated Cys26 at the N-terminus. The N-termini of molecules *A* and *D* are on the same side of the tetrameric assembly but are opposite to those of molecules *B* and *C*. As explained, there are no direct interactions formed between molecules *A* and *D* or molecules *B* and *C*. That the bacterial two-hybrid experiments reveal a propensity for self-interaction of the truncated protein in the cytoplasm is in one sense consistent with the crystal structure of the truncated version of *Sm*Lip, which shows a tetrameric assembly containing a plausible dimer. On the other hand, the structure of the tetramer is incompatible with dimeric or tetrameric

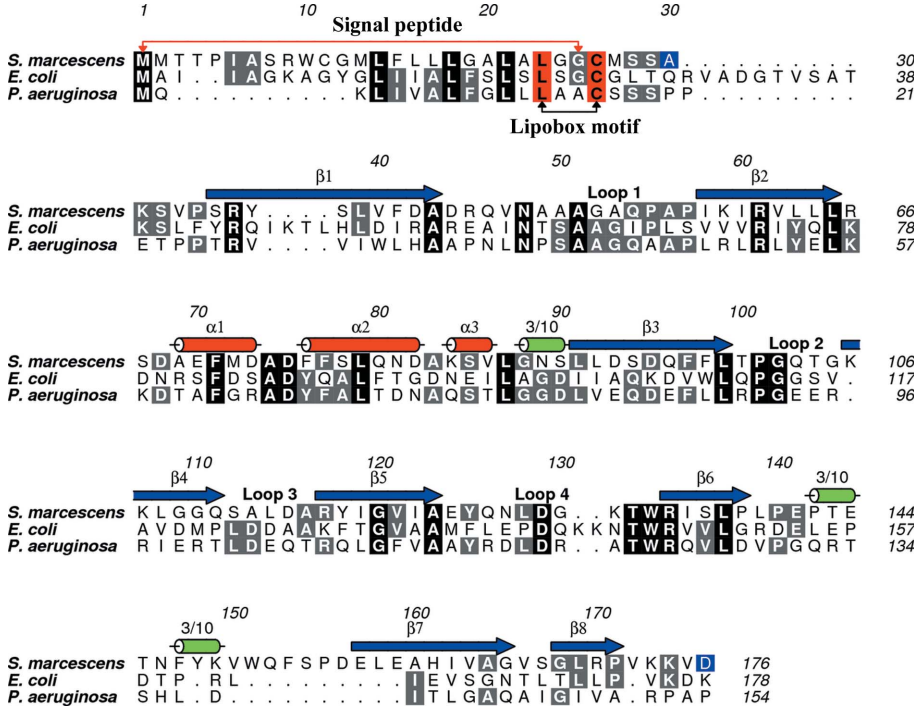


Figure 4
The primary and secondary structure of *Sm*Lip and sequence alignment with two homologues. *S. marcescens* Lip is aligned with the homologous proteins from enteroaggregative *E. coli* (GenBank CBG37366.1) and *P. aeruginosa* (NCBI Reference Sequence NP_248770.1, PA0080). The secondary structure of *Sm*Lip is depicted with blue arrows for β-strands and red and green cylinders for α-helices and 3_{10}-helices, respectively. Residues conserved in all three sequences are shown in black boxes and those conserved in only two sequences are shown in grey boxes. The start and finish of the lipobox motif are marked by red boxes; the residues at the start and end of the sequence used in the structure analysis (Ala30–Asp176) are shown in blue boxes. The alignment was generated using *T-Coffee* (Di Tommaso *et al.*, 2011) in the M-Coffee mode and the figure was prepared using *ALINE* (Bond & Schüttelkopf, 2009).

structures if the N-terminus is membrane-bound. These observations may be a result of the different concentrations and experimental conditions used. We suggest that *Sm*Lip is a membrane-bound monomer but displays a propensity to interact with itself.

A reviewer commented on the possibility that the *Sm*Lip tetramer might represent an inactive or alternative state of the protein. This is an intriguing suggestion and raises questions about how conversion to an active form might occur and how the T6SS itself is regulated. We have no data to address this issue and further studies would be required to investigate such a possibility.

3.5. Comparisons with structural homologues

A search for structural neighbours in the Protein Data Bank using the *PDBeFold* (http://pdbe.org/fold) and *ProFunc* servers (Laskowski *et al.*, 2005) gives a *Z* score of 6.1 with sea bream transthyretin (Eneqvist *et al.*, 2004; PDB entry 1sn0). This matched 84 residues with an r.m.s.d. of 2.7 Å. The β-sheet structures align well (Supplementary Fig. S3). The r.m.s.d. and relatively low *Z* score reflect the low sequence identity shared between the two proteins of approximately 7%. Nevertheless, the structural relationship is clear and *Sm*Lip can be classed as a new member of the transthyretin-like protein family. The only other member of this protein family is 5-hydroxyisourate hydrolase (EC 3.5.2.17; Hennebry *et al.*, 2006), an enzyme that is found only in prokaryotes, leading to the conclusion that this represents an example of divergent evolution (Hennebry, 2009). The sequence identity shared between this hydrolase and *Sm*Lip is only 6%, but the similarity in fold is evident (data not shown). We carried out further comparisons seeking to inform on Lip function.

Transthyretin binds the hormone thyroxine, self-interacts to form a tetramer and also forms a complex with retinol-binding protein (Blake *et al.*, 1978; Wojtczak *et al.*, 1992; Monaco *et al.*, 1995; Zanotti *et al.*, 2008). In common with transthyretin, *Sm*Lip forms a tetrameric assembly. However, the *Sm*Lip oligomer is distinct and an overlay of one *Sm*Lip polypeptide with a subunit from transthyretin does not produce an overlap of any of the other polypeptides (data not shown).

Transthyretin forms a dimer by antiparallel self-association of the β6 and β8 strands, creating a curved eight-stranded β-sheet (Blake *et al.*, 1978). The binding of the hormone thyroxine occurs at the tetramer interface created by the convex surfaces of two of these eight-stranded β-sheets as the protein assembles as a dimer of dimers. The thyroxine-binding residues in transthyretin are not conserved in *Sm*Lip and an overlay of an *Sm*Lip polypeptide and transthyretin subunit places the ligand-binding site on the surface of the former (Supplementary Fig. S3). It is unlikely that *Sm*Lip acts to bind hydrophobic ligands of the type that transthyretin can bind.

Transthyretin associates with retinol-binding protein using residues in three turns: two from one subunit that link β1 to β2 and β4 to β5, and one from another subunit that links β1 to β2 (Monaco *et al.*, 1995). These parts of the transthyretin structure correspond to loops 1 and 3 of *Sm*Lip. Loop 1 is directed out from the globular fold into the periplasmic space; it is placed to interact with physiological partners and may represent a binding site for other proteins/molecules.

In a recent study of *Klebsiella pneumoniae* 5-hydroxyisourate hydrolase, the residues important for catalytic function were confirmed as His7, Arg41, His92 and Ser108, which together with Tyr105 form a polar and symmetric active site at a dimer interface (French & Ealick, 2011). A structure-based sequence alignment matches four of these catalytic residues (with the exception being Ser108) to Asp42, Gly105, His161 and Val172, respectively, in *Sm*Lip. The polypeptides do not overlay in the vicinity of Ser108 (data not shown) and it is unlikely that Lip has any hydrolase activity.

The biological role of *Sm*Lip or its orthologues in the T6SS has yet to be unambiguously defined. Structural comparisons appear to rule out, rather than assign, a function. The propensity to self-associate using parts of the *Sm*Lip structure that will be exposed in the periplasm suggests that this protein, exploiting the lipid anchor, helps to bind and position different components of the secretion apparatus at the outer membrane. Future experiments, aided by the structural model, can address this hypothesis.

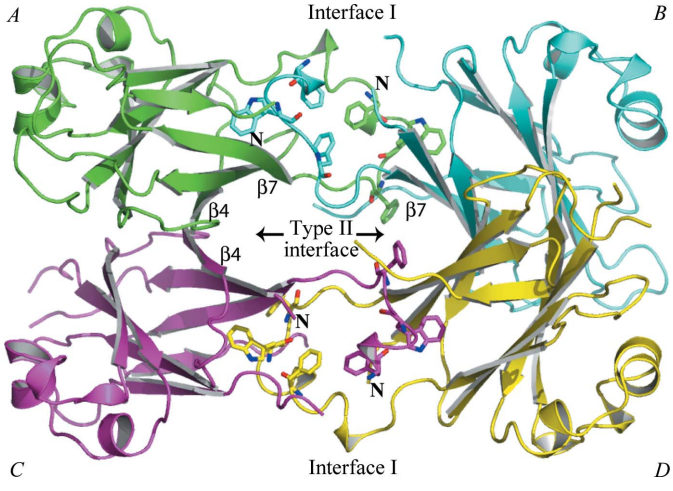


Figure 5
The asymmetric unit. The four molecules that constitute the asymmetric unit are shown in different colours using the secondary-structure assignment given in Fig. 1 and labelled. The two types of protein–protein interface are labelled, as are the β4 and β7 strands. Residues that contribute significantly to the type I interface (Phe147, Trp151 and Phe153) are depicted as sticks.

This work was funded by the European Commission Seventh Framework Programme (FP7/2007-2013), the Aeropath project, the Wellcome Trust (grants 082596 and 083481 and a PhD studentship to GE) and a Royal Society Project Grant and a Royal Society of Edinburgh Personal Research Fellowship (SJC). We acknowledge the contribution of the Pathogen Sequencing Unit at the Wellcome Trust Sanger Institute, Hinxton, England for performing the *S. marcescens* Db11 genome-sequencing project.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.

Aschtgen, M. S., Bernard, C. S., De Bentzmann, S., Lloubès, R. & Cascales, E. (2008). *J. Bacteriol.* **190**, 7523–7531.

Blake, C. C., Geisow, M. J., Oatley, S. J., Rérat, B. & Rérat, C. (1978). *J. Mol. Biol.* **121**, 339–356.

Blondel, C. J., Yang, H.-J., Castro, B., Chiang, S., Toro, C. S., Zaldívar, M., Contreras, I., Andrews-Polymenis, H. L. & Santiviago, C. A. (2010). *PLoS One*, **5**, e11724.

Bond, C. S. & Schüttelkopf, A. W. (2009). *Acta Cryst.* D**65**, 510–512.

Bönemann, G., Pietrosiuk, A. & Mogk, A. (2010). *Mol. Microbiol.* **76**, 815–821.

Bos, M. P., Robert, V. & Tommassen, J. (2007). *Annu. Rev. Microbiol.* **61**, 191–214.

Cascales, E. (2008). *EMBO Rep.* **9**, 735–741.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F. & Notredame, C. (2011). *Nucleic Acids Res.* **39**, W13–W17.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Eneqvist, T., Lundberg, E., Karlsson, A., Huang, S., Santos, C. R., Power, D. M. & Sauer-Eriksson, A. E. (2004). *J. Biol. Chem.* **279**, 26411–26416.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Filloux, A., Hachani, A. & Bleves, S. (2008). *Microbiology*, **154**, 1570–1583.

French, J. B. & Ealick, S. E. (2011). *Acta Cryst.* D**67**, 671–677.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.

Gerlach, R. G. & Hensel, M. (2007). *Int. J. Med. Microbiol.* **297**, 401–415.

Hatzixanthis, K., Palmer, T. & Sargent, F. (2003). *Mol. Microbiol.* **49**, 1377–1390.

Hejazi, A. & Falkiner, F. R. (1997). *J. Med. Microbiol.* **46**, 903–912.

Hennebry, S. C. (2009). *FEBS J.* **276**, 5367–5379.

Hennebry, S. C., Law, R. H., Richardson, S. J., Buckle, A. M. & Whisstock, J. C. (2006). *J. Mol. Biol.* **359**, 1389–1399.

Holland, I. B. (2010). *Methods Mol. Biol.* **619**, 1–20.

Hood, R. D. *et al.* (2010). *Cell Host Microbe*, **7**, 25–37.

Jani, A. J. & Cotter, P. A. (2010). *Cell Host Microbe*, **8**, 2–6.

Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. & Krogh, A. (2003). *Protein Sci.* **12**, 1652–1662.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 5752–5756.

Karimova, G., Ullmann, A. & Ladant, D. (2000). *Methods Enzymol.* **328**, 59–73.

Karimova, G., Ullmann, A. & Ladant, D. (2001). *J. Mol. Microbiol. Biotechnol.* **3**, 73–82.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Laskowski, R. A., Watson, J. D. & Thornton, J. M. (2005). *Nucleic Acids Res.* **33**, W89–W93.

Liu, H., Coulthurst, S. J., Pritchard, L., Hedley, P. E., Ravensdale, M., Humphris, S., Burr, T., Takle, G., Brurberg, M. B., Birch, P. R., Salmond, G. P. & Toth, I. K. (2008). *PLoS Pathog.* **4**, e1000093.

MacIntyre, D. L., Miyata, S. T., Kitaoka, M. & Pukatzki, S. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 19520–19524.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Monaco, H. L., Rizzi, M. & Coda, A. (1995). *Science*, **268**, 1039–1041.

Murdoch, S. L., Trunk, K., English, G., Fritsch, M. J., Pourkarimi, E. & Coulthurst, S. J. (2011). *J. Bacteriol.* **193**, 6057–6069.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Pukatzki, S., McAuley, S. B. & Miyata, S. T. (2009). *Curr. Opin. Microbiol.* **12**, 11–17.

Sarris, P. F., Skandalis, N., Kokkinidis, M. & Panopoulos, N. J. (2010). *Mol. Plant Pathol.* **11**, 795–804.

Schwarz, S., Hood, R. D. & Mougous, J. D. (2010). *Trends Microbiol.* **18**, 531–537.

Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Wojtczak, A., Luft, J. & Cody, V. (1992). *J. Biol. Chem.* **267**, 353–357.

Zanotti, G., Folli, C., Cendron, L., Alfieri, B., Nishida, S. K., Gliubich, F., Pasquato, N., Negro, A. & Berni, R. (2008). *FEBS J.* **275**, 5841–5854.

APPENDIX C

English G, Trunk K, Rao VA, Srikannathasan V, Hunter WN and Coulthurst SJ (2012) New secreted toxins and immunity proteins encoded within the Type VI secretion system gene cluster of *Serratia marcescens*. ***Mol Microbiol*** **doi: 10.1111/mmi.12028**

 doi:10.1111/mmi.12028

New secreted toxins and immunity proteins encoded within the Type VI secretion system gene cluster of *Serratia marcescens*

Grant English,[1†] Katharina Trunk,[1†] Vincenzo A. Rao,[2] Velupillai Srikannathasan,[2] William N. Hunter[2] and Sarah J. Coulthurst[1*]
[1]*Division of Molecular Microbiology, College of Life Sciences, University of Dundee, Dundee, UK.*
[2]*Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee, UK.*

Summary

Protein secretion systems are critical to bacterial virulence and interactions with other organisms. The Type VI secretion system (T6SS) is found in many bacterial species and is used to target either eukaryotic cells or competitor bacteria. However, T6SS-secreted proteins have proven surprisingly elusive. Here, we identified two secreted substrates of the antibacterial T6SS from the opportunistic human pathogen, *Serratia marcescens*. Ssp1 and Ssp2, both encoded within the T6SS gene cluster, were confirmed as antibacterial toxins delivered by the T6SS. Four related proteins encoded around the Ssp proteins ('Rap' proteins) included two specifically conferring self-resistance ('immunity') against T6SS-dependent Ssp1 or Ssp2 toxicity. Biochemical characterization revealed specific, tight binding between cognate Ssp–Rap pairs, forming complexes of 2:2 stoichiometry. The atomic structures of two Rap proteins were solved, revealing a novel helical fold, dependent on a structural disulphide bond, a structural feature consistent with their functional localization. Homologues of the *Serratia* Ssp and Rap proteins are found encoded together within other T6SS gene clusters, thus they represent founder members of new families of T6SS-secreted and cognate immunity proteins. We suggest that Ssp proteins are the original substrates of the *S. marcescens* T6SS, before horizontal acquisition of other T6SS-secreted toxins. Molecular insight has been provided into how pathogens utilize antibacterial T6SSs to overcome competitors and succeed in polymicrobial niches.

Accepted 3 September, 2012. *For correspondence. E-mail s.j.coulthurst@dundee.ac.uk; Tel. (+44) 1382 386208; Fax (+44) 1382 388216. †These authors contributed equally and are listed alphabetically.

Introduction

Protein secretion systems and their substrates are central to bacterial virulence and interaction with other organisms (Gerlach and Hensel, 2007). Six different secretion systems (Types I–VI) are used by Gram-negative bacteria to transport specific proteins to the exterior of the bacterial cell or further inject them into target cells. The most recently described of these is the Type VI secretion system (T6SS) (Filloux *et al.*, 2008). T6SSs are complex multi-protein assemblies that span both bacterial membranes and inject effector proteins directly from the bacterial cytoplasm into target cells (Bonemann *et al.*, 2010; Cascales and Cambillau, 2012). T6SSs are encoded by large, variable gene clusters that contain 13 'core' essential components, believed to make up the basic secretion apparatus. Two core proteins, Hcp and VgrG, form the extracellular part of the secretion machinery and depend on a functional T6SS apparatus for their movement to the outside of the bacterial cell (indeed, the presence of Hcp in the secreted fraction has provided a useful assay for basic T6SS assembly and activity; Pukatzki *et al.*, 2009). Hcp and VgrG most likely form a needle-like membrane-puncturing device related to the bacteriophage tail spike; this structure is believed to be pushed to the outside of the secreting cell and likely into target cells upon contraction of a tail sheath-like structure (Leiman *et al.*, 2009; Bonemann *et al.*, 2010; Basler *et al.*, 2012). T6SSs occur in many pathogenic bacteria and are implicated in virulence in important pathogens, including *Burkholderia mallei*, *Burkholderia pseudomallei*, *Burkholderia cenocepacia*, *Vibrio cholerae*, *Aeromonas hydrophila*, *Edwardsiella tarda* and *Pseudomonas aeruginosa* (Zheng and Leung, 2007; Cascales, 2008; Jani and Cotter, 2010; de Pace *et al.*, 2010; Burtnick *et al.*, 2011; Rosales-Reyes *et al.*, 2012). In several cases, the action of such 'anti-eukaryotic' T6SSs appears to result in disruption of the actin cytoskeleton (Pukatzki *et al.*, 2007; Aubert *et al.*, 2008; Suarez *et al.*, 2010). Exciting recent work has dem-

onstrated that some T6SSs are used to target other bacteria, efficiently killing or inhibiting competitors. This has been reported for T6SSs in *P. aeruginosa*, *Burkholderia thailandensis*, *V. cholerae* and *Serratia marcescens* (Hood *et al*., 2010; MacIntyre *et al*., 2010; Schwarz *et al*., 2010; Murdoch *et al*., 2011). The discovery that certain T6SSs may be 'antibacterial' rather than, or in addition to, 'anti-eukaryotic' is highly relevant to the competitive fitness and success of pathogens, particularly within polymicrobial infection sites. Such a system could provide the pathogen with a large competitive advantage against other bacteria in the host or the environment, enabling it to proliferate and mount a successful infection.

Identifying the proteins secreted by T6SSs is a priority, as they will be the 'effectors' that directly act on target eukaryotic or bacterial cells. A special case of T6-secreted effector is the class of 'evolved' VgrG proteins found in a minority of T6SSs which have extra C-terminal effector domains, e.g. the actin cross-linking domain of *V. cholerae* VgrG1 that is translocated into mammalian cells (Pukatzki *et al*., 2007; Jani and Cotter, 2010). However, excluding the structural components VgrG and Hcp, very few 'true' T6SS-secreted proteins have been confirmed so far. Best characterized are three effector proteins, antibacterial toxins named Tse1–3, secreted by the antibacterial HSI-1 T6SS of *P. aeruginosa*. Tse1 and 3 are peptidoglycan hydrolases that attack the cell wall of target bacteria (Hood *et al*., 2010; Russell *et al*., 2011). Tse2 is active in the cytoplasm of target cells, where it efficiently induces quiescence (Li *et al*., 2012). All three have adjacently encoded cognate 'immunity' proteins (Tsi1–Tsi3) which protect the secreting cell from harming itself or being harmed by its sibling neighbours (Hood *et al*., 2010; Russell *et al*., 2011). Significantly, obvious homologues of the Tse and Tsi proteins are not detectable outside of *P. aeruginosa* (Hood *et al*., 2010). A recent report has also identified a number of candidate T6SS substrates in *B. thailandensis*, one of which was confirmed as a new T6-secreted peptidoglycan amidase (Russell *et al*., 2012).

Opportunistic Gram-negative bacteria cause a large proportion of problematic and antibiotic-resistant hospital-acquired infections. Enterobacteria (especially extended-spectrum β-lactamase producing isolates) are among the leading culprits, including *S. marcescens* (Choi *et al*., 2007; Lockhart *et al*., 2007). We previously reported that *S. marcescens* Db10 possesses a T6SS with potent antibacterial activity (Murdoch *et al*., 2011). How this activity is mediated, in particular the antibacterial effectors secreted by the T6SS, was unknown. Hence, we sought to identify and characterize novel antibacterial effectors secreted by the *S. marcescens* T6SS. We report the identification and characterization of two such effectors, Ssp1 and Ssp2, which are encoded within the T6SS gene cluster and represent novel T6-secreted antibacterial toxins. We have also identified and characterized the Rap proteins, which include the cognate immunity proteins to these toxins. Biochemical analyses demonstrated a tight and specific interaction between secreted and immunity proteins. These secreted toxins and immunity proteins represent two new protein families, co-occurring within T6SS gene clusters of many other organisms. Additionally, determination of high-resolution crystal structures of two members of the Rap protein family revealed that this family possesses a previously undescribed protein fold that is dependent on formation of a disulphide bond.

Results

The T6SS gene cluster harbours self-resistance determinants and candidate secreted effectors

The T6SS gene cluster of *S. marcescens* Db10, *SMA2244–2281*, contains 38 genes, including many with no known function (Murdoch *et al*., 2011). We speculated that encoded within this cluster might be T6-secreted effectors and/or self-resistance determinants, the latter preventing the T6SS-expressing cell from harming itself or being harmed by its isogenic (sibling) neighbours. In order to determine whether the cluster did indeed contain self-resistance determinants (such as specific immunity proteins analogous to the Tsi proteins), we generated a mutant lacking the entire T6SS gene cluster (ΔT6SS) and examined whether it was fully resistant to the T6SS of the wild type strain. Co-culture of two target strains, wild type Db10 and the ΔT6SS mutant, each with a wild type and a Δ*clpV* attacker, showed that the ΔT6SS mutant had lost resistance to T6SS-mediated inhibition or killing by the wild type strain (Fig. 1A, left). Recovery of ΔT6SS was decreased 100-fold when it was co-cultured with the wild type strain, compared with when the wild type strain was co-cultured with itself. This effect was dependent on a functional T6SS in the attacker as there was no loss of ΔT6SS when it was co-cultured with a Δ*clpV* mutant. The ATPase ClpV is one of the core, structural components of the T6SS and we have shown previously that it is essential for Hcp secretion and T6-mediated antibacterial killing activity of *S. marcescens* Db10 (Murdoch *et al*., 2011). To simplify the analysis of multiple mutants, we defined the 'resistance index' of a strain as the difference between recovery when co-cultured with the wild type strain and recovery when co-cultured with the Δ*clpV* mutant, specifically $\log_2$[recovery vs. wild type/recovery vs. Δ*clpV*]. The wild type or other resistant strain will show no difference and have a resistance index of 0. A target strain with reduced ability to resist harm caused by the T6SS will have a negative resistance index, exemplified by the ΔT6SS mutant with a resistance index of −7.6 (Fig. 1A, right). Self-resistance did not depend on an active T6SS,

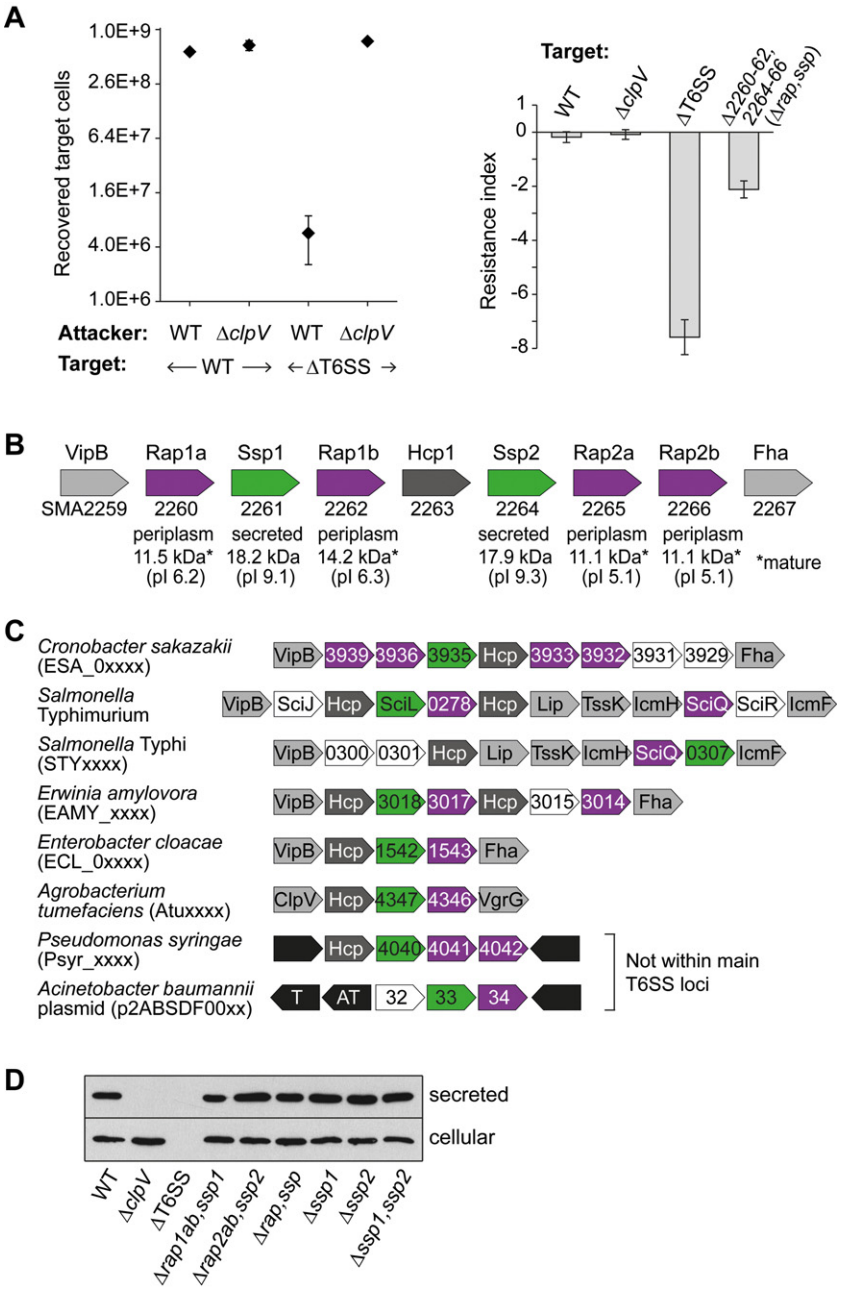


Fig. 1. An internal locus in the *S. marcescens* T6SS gene cluster encodes secreted proteins and self-resistance functions.
A. T6SS-mediated inhibition of self in the absence of genes within the T6SS gene cluster. Left: number of recovered target cells, either wild type (WT) or ΔT6SS mutant (Δ*SMA2244–2281*), following co-culture (1:1) with the attacking strain, WT or Δ*clpV* (T6SS inactive mutant). Right: resistance index, defined as $\log_2$[recovery of target in presence of wild type/recovery of target in presence of Δ*clpV*], of wild type Db10, Δ*clpV*, ΔT6SS or a mutant lacking genes *SMA2260–2262* and *SMA2264–2266*. Bars show mean ± SEM ($n = 4$).
B. and C. Schematic depiction of loci containing genes encoding the Rap and Ssp genes in *S. marcescens* Db10 (B) and homologues in selected other organisms (C). Rap family proteins are shown in purple, Ssp family proteins in green, conserved T6SS core genes in grey, and Hcp homologues in dark grey. In (B) cellular localization and theoretical mass and pI of the proteins are given beneath the corresponding gene.
D. Secretion of Hcp1 by wild type and mutants of *S. marcescens* Db10 as shown by anti-Hcp1 immunoblotting of cellular and secreted fractions. Δ*rap1ab,ssp1* indicates a mutant lacking the *rap1a*, *rap1b* and *ssp1* genes; Δ*rap2ab,ssp2* indicates a mutant lacking the *rap2a*, *rap2b* and *ssp2* genes; Δ*ssp1,ssp2* indicates a mutant lacking the *ssp1* and *ssp2* genes, and Δ*rap,ssp* indicates a mutant lacking all of the *rap* and *ssp* genes.

as a Δ*clpV* mutant had a resistance index of 0. Similarly, mutants in other essential core T6SS components, Δ*lip*, Δ*icmH* and Δ*tssK* (Murdoch *et al.*, 2011) also showed no loss of self-resistance (data not shown). Therefore, genes other than those encoding the core conserved T6SS components were implicated in self-resistance. Our attention was particularly caught by a locus in the middle of the gene cluster, where six non-conserved genes, *SMA2260–2262*, *SMA2264–2266*, are flanked by conserved T6SS components. A mutant lacking all of these genes (but maintaining intact *SMA2263*, encoding Hcp1) was also tested and found to have a negative resistance index (Fig. 1A). Hence, one or more of these genes contributes to self-resistance and may encode immunity protein(s).

Closer examination of the proteins encoded by *SMA2260–2262* and *SMA2264–2266* revealed two classes of small proteins (Fig. 1B). SMA2261 and SMA2264 were basic proteins with detectable sequence similarity between them, no discernable cellular localization signals and no predicted function. We hypothesized that they might be secreted substrates, and, given subsequent results, named them Ssp1 and Ssp2 (Secreted small protein). SMA2260, SMA2262, SMA2265 and SMA2266 were proteins with classical Sec-dependent

N-terminal signal peptides, identified using SignalP (Petersen *et al.*, 2011), and thus predicted to be periplasmic. They also had no predicted function but shared detectable sequence similarity with each other. We hypothesized that they would represent specific immunity proteins or other self-resistance determinants, and named them Rap1a, Rap1b, Rap2a and Rap2b (Resistance associated protein). The genes encoding all these small proteins fall into two sets, either side of *hcp1* and within genes encoding T6SS structural proteins, *vipB* and *fha* (Fig. 1B). Homologues of the Ssp and Rap proteins are encoded within T6SS gene clusters in many other bacterial species (and in at least one case apart from the T6SS) and appear to always co-occur (see representative examples in Fig. 1C). We speculated that the *S. marcescens* Ssp and Rap proteins represented previously unknown combinations of T6-secreted effectors and cognate immunity proteins. Importantly, mutants lacking one, some or all of the small proteins exhibited wild type levels of Hcp secretion (Figs 1D and S1). Thus, none of the Ssp or Rap proteins is required for Hcp secretion, i.e. they play no structural role in the T6SS.

Ssp1 *and* Ssp2 *are Type VI-secreted effectors*

Ssp1 and Ssp2 were shown to be secreted substrates of the T6SS by immunoblotting secreted fractions from the wild type strain, two T6SS mutants, Δ*clpV* and Δ*tssE*, and the corresponding complemented strains, using specific anti-Ssp1 and anti-Ssp2 antibodies. Both proteins were detected in the culture supernatant, entirely dependent on a functional T6SS (Fig. 2A). Neither protein was detectable in the cellular fraction, implying they were rapidly turned over if not secreted (data not shown). Secretion of Ssp2 was also independent of Ssp1, and vice versa, again confirming they have no structural or accessory role in the secretion machinery (Fig. 2A). In order to establish if the Ssp proteins were antibacterial toxins contributing to the killing or inhibition of the susceptible ΔT6SS mutant by the wild type strain, the recovery of ΔT6SS in the presence of different mutants was determined. Recovery of the ΔT6SS target was increased fivefold with the Δ*ssp2* mutant, or any multiple mutant lacking *ssp2*, as attacker, compared with the wild type attacker (Fig. 2B). In contrast, in this assay, loss of Ssp1 did not cause a statistically significant impairment in killing of ΔT6SS. None of the Rap proteins was required for targeting of ΔT6SS (Fig. 2B).

Self-resistance against the Ssp1 *and* Ssp2 *toxins is mediated by their cognate* Rap *partners*

In order to assess the contribution of each of the *rap* (and *ssp*) genes to self-resistance, single mutants were constructed. It rapidly became apparent that single mutants in *rap1a* and *rap2a* had severe fitness defects on solid media, both on rich (Fig. 3A) and minimal media (Fig. S2), and particularly Δ*rap2a*. Culture spots and single colonies of each mutant were 'thinner', smaller (in the case of Δ*rap2a*), and with altered surface morphology. Examination of single cells showed that the Δ*rap1a* cells appeared bigger and less uniformly shaped than wild type cells and the Δ*rap2a* cells had striking phenotypes, being either markedly distended or highly elongated (Figs 3A and S2). These phenotypes could be complemented by expression of the corresponding gene *in trans* (indeed, the additional stress of the selective antibiotic made the phenotypes of

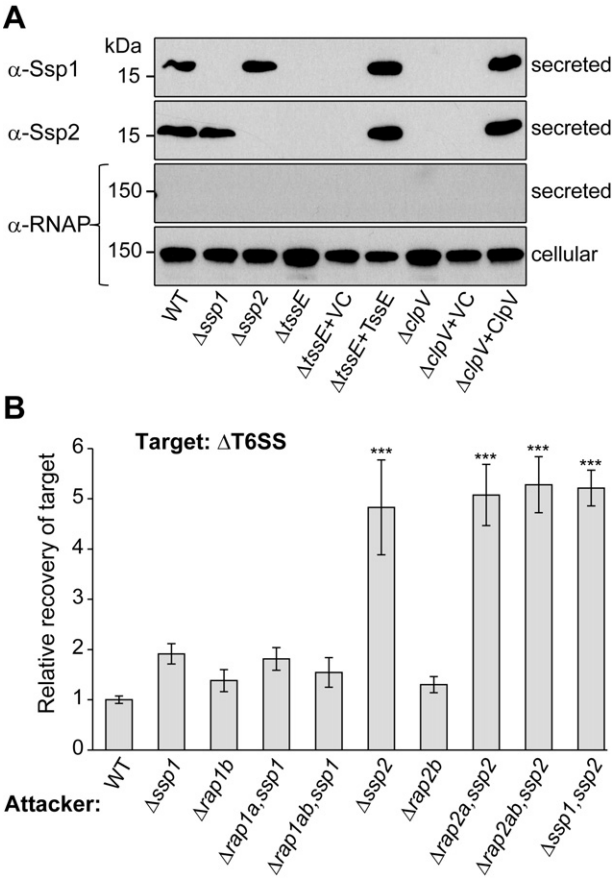


Fig. 2. Proteins Ssp1 and Ssp2 are secreted by the Type VI secretion system and self-harm is mediated by Ssp2.
A. Immunoblot detection of Ssp1 and Ssp2 in the secreted fraction of the strains indicated using antibodies against Ssp1, Ssp2 or RNAP (lysis control); levels of RNAP in the cellular fraction are also shown. Strains are: wild type *S. marcescens* Db10 (WT); mutants Δ*ssp1*, Δ*ssp2*, Δ*tssE*, Δ*clpV*; mutants carrying vector control plasmids (+VC, pSUPROM) and mutants carrying complementing plasmids (Δ*tssE* + TssE, pSC045; Δ*clpV* + ClpV, pSC039).
B. Recovery of the ΔT6SS mutant as the target strain following co-culture with the different attacking strains indicated, expressed relative to recovery of ΔT6SS when co-cultured with wild type Db10. Bars show mean ± SEM ($n \geq 4$); *** indicates a significant difference compared with the wild type strain ($P < 0.001$).

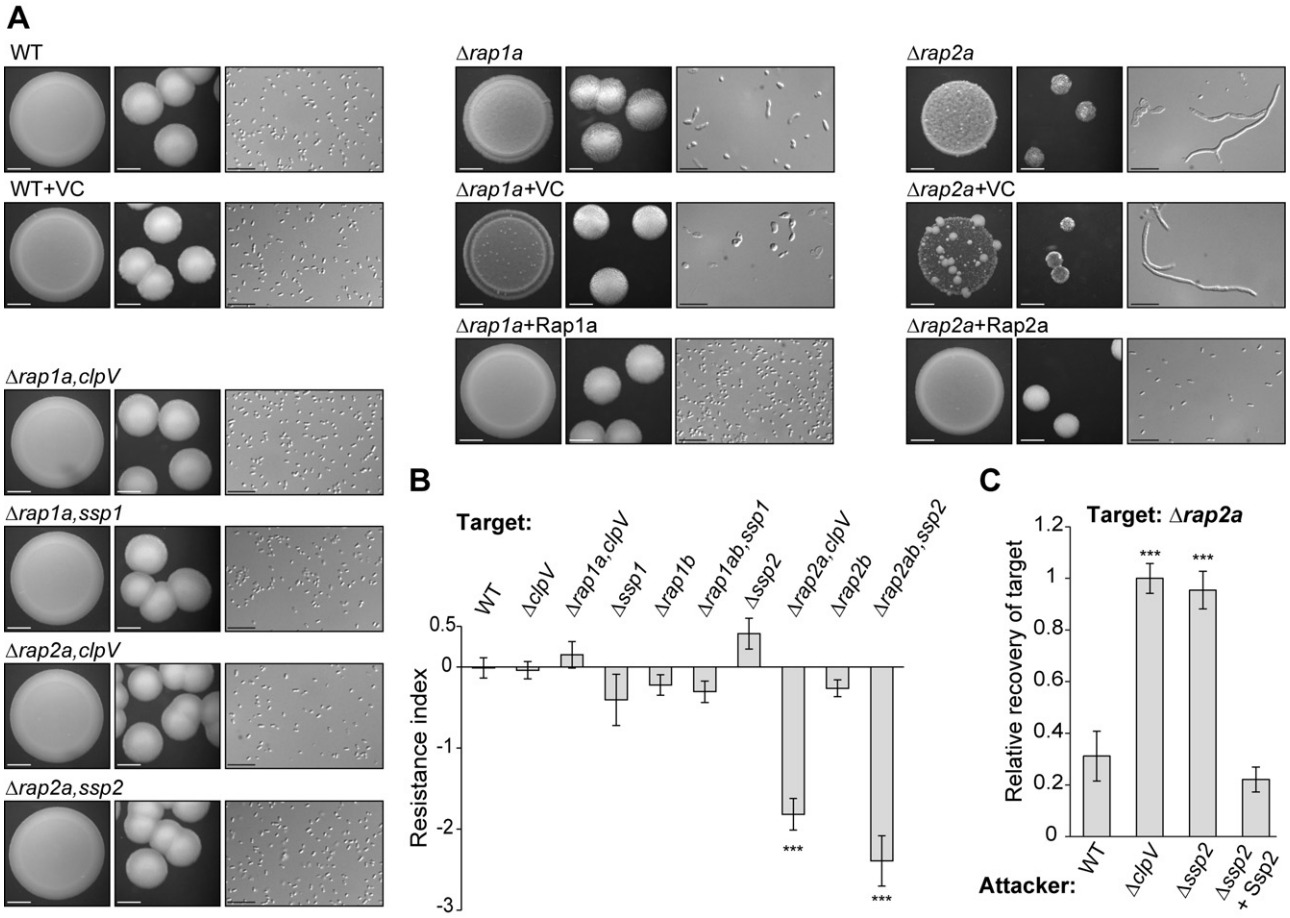


Fig. 3. Self-resistance is mediated by specific Rap immunity proteins cognate to the secreted Ssp proteins.
A. Phenotypes of wild type *S. marcescens* Db10 (WT) and selected single and double mutants after growth on solid LB media for 24 h. For each strain, representative images of the morphology of a culture spot (left, scale bar 2 mm), single colonies (middle, scale bar 1 mm) and individual cells (right, scale bar 10 μm) are shown. Mutants carrying complementing plasmids are Δ*rap1a* + Rap1a (pSC538) and Δ*rap2a* + Rap2a (pSC542); the vector control plasmid (VC) was pSUPROM.
B. Resistance index of wild type Db10 and the deletion mutants indicated as target strains.
C. Recovery of a Δ*rap2a* mutant (Δ*rap2a*,Δ*clpV*) as the target strain following co-culture with the different attacking strains indicated, expressed relative to recovery of target when co-cultured with the Δ*clpV* mutant. All strains carry the vector control plasmid pSUPROM, except for Δ*ssp2* + Ssp2, in which the mutant carries the complementing plasmid pSC541. In (B) and (C), bars show mean ± SEM ($n \geq 4$); *** indicates a significant difference compared with the wild type strain ($P < 0.001$).

Δ*rap1a* and Δ*rap2a* mutants carrying vector control plasmids even more pronounced; Fig. 3A). In contrast, when Δ*rap1a* was constructed in combination with Δ*ssp1*, or Δ*rap2a* in combination with Δ*ssp2*, the double mutants were readily made and were of normal appearance and fitness on solid media (Figs 3A and S2). These observations are entirely consistent with Rap2a serving as the immunity protein against the Ssp2 toxin and Rap1a being the cognate immunity protein alleviating toxicity mediated by Ssp1. Similarly, when Δ*rap1a* or Δ*rap2a* mutants were constructed in combination with the Δ*clpV* mutation, the double mutants were again apparently healthy on solid media (Figs 3A and S2). Hence, self-toxicity depended on a functional T6SS, implying it was caused by T6SS-mediated injection of the toxin into a susceptible cell by its neighbours (further supported by the observation that Δ*rap1a* and Δ*rap2a* did not display comparable growth defects in liquid culture, where contact-dependent targeting is unlikely to occur efficiently; Fig. S2).

An immunity function for Rap2a was directly demonstrated in co-culture (antibacterial competition assay): target strains containing a Δ*rap2a* mutation all showed a negative resistance index (Fig. 3B; the single Δ*rap2a* mutant was not tested because of its severe sickness). Strains containing a Δ*rap1a* mutation did not have a negative resistance index, consistent with the lack of significant contribution of Ssp1 to self-killing under the conditions of these assays (Fig. 2B). To confirm that Ssp2 was directly and entirely responsible for the inhibitory effect of a wild type attacker strain on Δ*rap2a*, we showed that a Δ*rap2a*

target strain is completely resistant to attack by a Δ*ssp2* mutant, with recovery of Δ*rap2a* in the presence of a Δ*ssp2* attacker being the same as its recovery in the presence of a Δ*clpV* attacker (Fig. 3C). Additionally, inhibition of Δ*rap2a* target was restored to wild type levels when the Δ*ssp2* mutant attacker was complemented by expression of Ssp2 *in trans*. The susceptibility of Δ*rap2a* to killing by the wild type strain could also be fully complemented (Fig. S2). Hence, Ssp2 and Rap2a represent a specific, cognate toxin and immunity protein pair.

Biochemical studies confirm a strong interaction between cognate secreted toxins and immunity proteins

Each of the Ssp and Rap proteins was overproduced and purified (in the case of the Rap proteins, without their N-terminal signal peptides). Complex formation was demonstrated using size exclusion chromatography (SEC) analysis. All Rap proteins were dimeric in solution, whereas Ssp1 and Ssp2 were monomeric (Fig. 4A and B). When equimolar amounts of Ssp1 and Rap1a were mixed, all of the protein was detected in a higher molecular weight complex (Fig. 4A) and the same was also observed on mixing Ssp2 and Rap2a (Fig. 4A). In contrast, no complex formation between Ssp1 and Rap1b or between Ssp2 and Rap2b, no additional three-way complexes and no 'cross' interactions between Ssp2 and Rap1a or between Ssp1 and Rap2a were observed (Figs 4B and S3). As mixing equimolar amounts of Ssp1 and Rap1a resulted in detection of only the complexed species, with no unbound form of either protein detectable, this implied a 1:1 molar complex. The 1:1 molar stoichiometry of the Rap1a–Ssp1 complex was confirmed by quantitative analysis of the composition of the higher molecular weight peak using in-gel SYPRO Orange staining (Rickman *et al*., 2004) (Fig. 4C). Given the dimeric nature of Rap1a and an apparent complex mass by SEC of around 55 kDa [predicted M_w of 1:1 molar complexes are 34.6 kDa (1:1), 69.2 kDa (2:2), 104 kDa (3:3) or 138 kDa (4:4)], this is most consistent with a heterotetrameric $Rap1a_2$–$Ssp1_2$ complex. Similar logic supported a $Rap2a_2$–$Ssp2_2$ complex (Fig. 4A and C). The SEC analyses suggested stable complexes were formed between the cognate Ssp and Rap proteins. Isothermal titration calorimetry (ITC) analysis of complex formation between Ssp1 and the Rap1a dimer and between Ssp2 and the Rap2a dimer (Fig. 4D) showed that, in both cases, binding was clearly exothermic and tight, at least in the low nanomolar K_d range. Although apparent K_d values of 4–8 nM were obtained, this is approaching the limit of accuracy for conventional ITC (Wiseman *et al*., 1989). Finally, we took advantage of the tight Rap–Ssp interactions to utilize immobilized Rap2a and Rap1a to affinity-purify secreted Ssp2 or Ssp1, respectively, from culture supernatant (Figs 4E and S3). Mass spectrometry not only confirmed the identity of Ssp1 and Ssp2 but also showed that Ssp2 is not processed on secretion, as almost complete sequence coverage revealed intact N- and C-termini (Fig. S3). For Ssp1, we observed an intact C-terminus but the sequence of the very N-terminus corresponds to several tryptic peptides too small to be detectable by standard mass spectrometry, so we were unable to definitively confirm lack of processing at this end (data not shown).

The secreted small proteins, Ssp1 *and* Ssp2*, exert distinct harmful effects when targeted to the periplasm in* Escherichia coli *and are neutralized by coexpression of the cognate immunity protein*

In order to confirm the antibacterial toxin function of Ssp1 and Ssp2 and establish in which cellular compartment they exerted their effect, each protein was produced in *E. coli*, either in the cytoplasm or artificially targeted to the periplasm (by the N-terminal fusion of the Sec-dependent OmpA signal peptide: sp-Ssp2 or sp-Ssp1). The presence of Ssp2 in the periplasm, but not the cytoplasm, prevented growth of *E. coli* on LB and minimal media; this toxicity was alleviated by the co-production of Rap2ab (Fig. 5A). Periplasmic Ssp1 was also toxic and its effect alleviated by co-production of Rap1ab (Fig. 5A). However Ssp1 toxicity was only observed at higher expression levels and on minimal media. The periplasmic localization of the Rap proteins in *S. marcescens* was confirmed by fractionation and immunoblotting of epitope-tagged versions of the protein (Fig. 5B, data not shown).

Next, we also determined whether Ssp2 and/or Ssp1 were essential for the observed antibacterial activity of the *S. marcescens* T6SS against other bacterial species (Murdoch *et al*., 2011). Neither protein was required for T6SS-dependent killing of other bacteria (Fig. S4). This implies redundancy of effector function, in other words that other T6SS-secreted toxins are still able to cause killing in the absence of Ssp1 and Ssp2.

The Rap *family of proteins display a new fold*

Bioinformatic analysis of the Rap protein sequences predicted similarities between them, despite relatively low sequence conservation, e.g. mature Rap1b and Rap2b share around 20% identity, as do mature Rap2a and Rap2b. All four Rap proteins appeared to be highly acidic, to carry an N-terminal periplasmic targeting signal sequence of about 24 residues, and were predicted to have similar α-helical structures (using *PSIPRED*). In addition, a cysteine pairing appeared to be conserved. Intriguingly, no structural relative of the Rap proteins could be identified; hence, we sought to obtain three-dimensional models

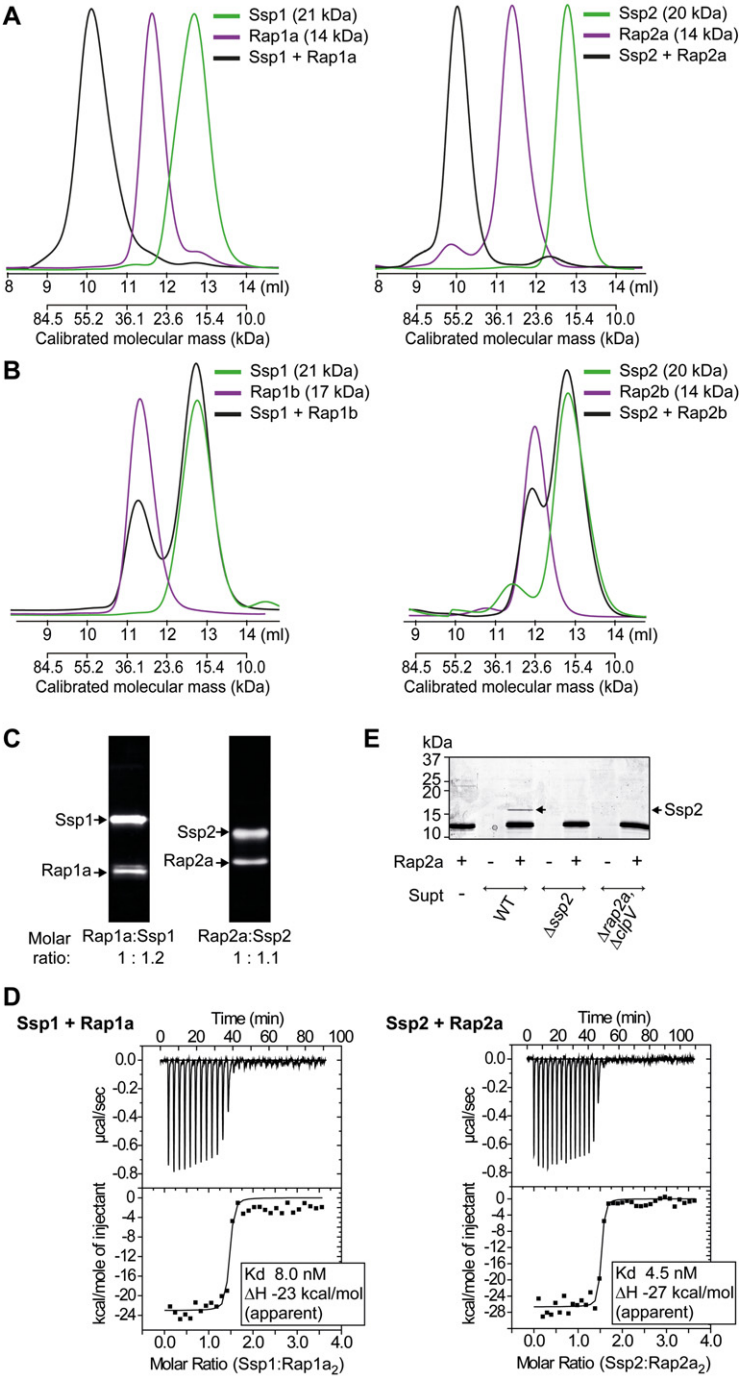


Fig. 4. *In vitro* interaction between cognate secreted toxins and immunity proteins.
A. and B. SEC analysis of complex formation between the proteins indicated. Ten nanomoles of the protein (or of each protein in the case of mixtures) was separated on a calibrated Superdex 75 10/300 GL column. The theoretical molecular mass of each monomer is given.
C. In-gel SYPRO Orange staining and quantification of the relative molar amounts of Rap1a and Ssp1 (left) or Rap2b and Ssp2 (right) in samples from the complex-containing peaks observed in SEC (A).
D. ITC analysis of the interaction between Ssp1 and Rap1a (left) and Ssp2 and Rap2a (right).
E. Affinity isolation of secreted Ssp2 from culture supernatant using immobilized immunity protein His-Rap2a as bait. Supernatant samples (Supt) were prepared from the strains indicated (WT, wild type). The band indicated by arrowhead was identified as Ssp2 by mass spectrometry.

using single crystal X-ray diffraction. Soluble recombinant Rap proteins were only obtained in high yield from *E. coli* Rosetta-gami (DE3), a strain engineered to support disulphide bond formation, and were dimeric in solution (Fig. 4). Well-ordered crystals of Rap1b and Rap2b were obtained and the structures were elucidated at 1.9 Å and 2.0 Å resolution respectively (Fig. 6, Table 1). A search for structural relatives using *PDBeFold* (pdbe.org/fold) and *ProFunc* (Laskowski *et al.*, 2005) was performed using monomers and dimers of Rap1b and Rap2b as templates. The only matches were to short segments of α-helices with Z-scores (< 3) indicating low statistical significance. The lack of any convincing structural relationships indicates that the fold observed in both Rap1b and Rap2b is previously uncharacterized.

The asymmetric unit of the Rap1b structure consists of a single subunit and crystallographic symmetry generates the dimer. Rap2b has two dimers, formed by subunits A : B

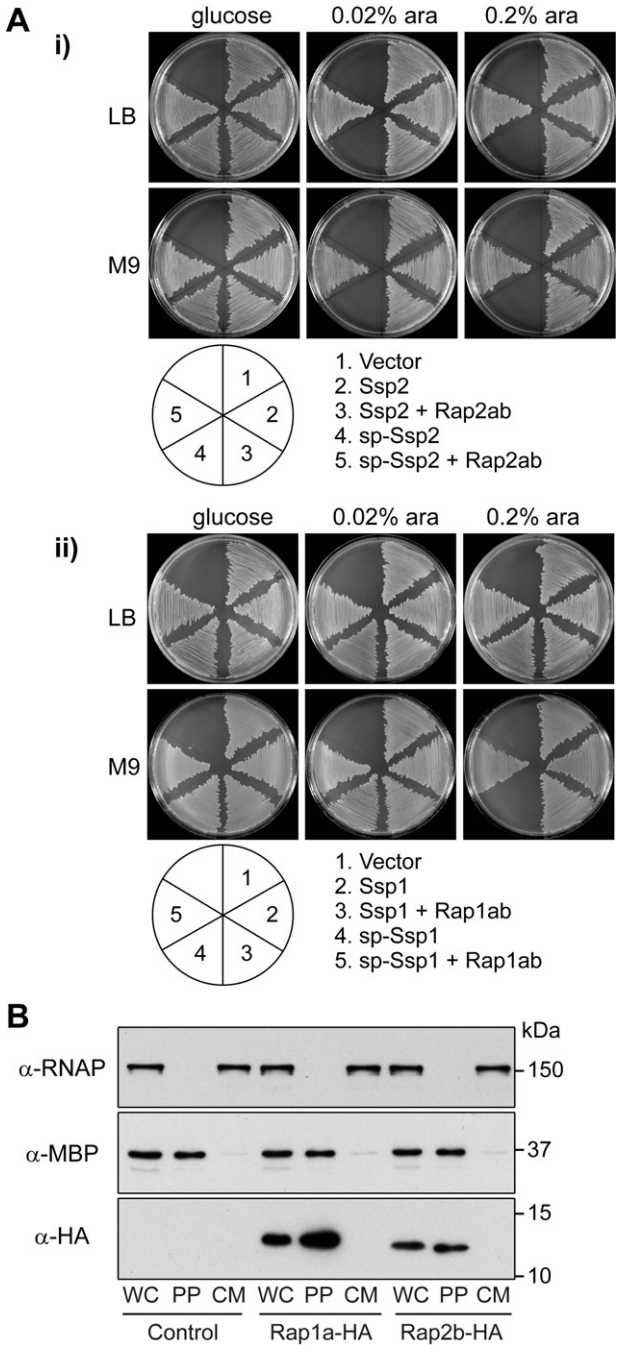


Fig. 5. Differential periplasmic toxicity of Ssp1 and Ssp2 and periplasmic localization of Rap proteins.
A. Growth of *E. coli* MG1655 transformed with (i) plasmids expressing Ssp2 (pSC133), Ssp2 + Rap2a + Rap2b (pSC134), $OmpA_{sp}$-Ssp2 (pSC138) or $OmpA_{sp}$-Ssp2 + Rap2a + Rap2b (pSC144) from an arabinose-inducible promoter, or with the empty vector (vector, pBAD18-Kn), on LB or M9 media containing 0.2% glucose, 0.02% arabinose or 0.2% arabinose; or (ii) the equivalent analysis using plasmids expressing Ssp1 (pSC151), Ssp1 + Rap1a + Rap1b (pSC159), $OmpA_{sp}$-Ssp1 (pSC152) or $OmpA_{sp}$-Ssp1 + Rap1a + Rap1b (pSC160).
B. Localization of RNAP (RNA polymerase, cytoplasmic marker protein), MBP (maltose binding protein, periplasmic marker protein) and HA-tagged Rap proteins. Wild type *S. marcescens* Db10 expressing Rap1a-HA (pSC538), Rap2b (pSC543) or the vector control (pSUPROM) was subjected to fractionation (WC, whole cell; PP, periplasm; CM, cytosol + membranes), followed by immunoblotting with anti-RNAP, anti-MBP or anti-HA antibodies as indicated.

and C : D, in the asymmetric unit. The four subunits are similar, with the root-mean-square deviation (r.m.s.d.) between superimposed Cα atoms ranging from 0.5 Å (subunit B and D) to 0.7 Å (subunits A and C). The two proteins display obvious similarities in terms of secondary, tertiary and quaternary structure (Figs 6 and S5). An overlay of Rap1b and Rap2b subunits matches 71 Cα atoms with a r.m.s.d. of 1.4 Å. Of note is the conservation of the disulphide bond, a key structural feature, likely critical to the stable folding of the subunit and subsequent dimerization of the Rap proteins. The Rap subunit is constructed around a helical bundle of five helices (α2–α6). A disulphide linkage (Cys54–Cys120 in Rap1b, Cys41–Cys101 in Rap2b) tethers α2 and α5 together. This association in turn supports formation of a hydrophobic core that is primarily aliphatic with residues contributed from α2, α4, α5 and α6. Three helices (α2, α3, α5) form a concave surface on one side of the monomer. A series of hydrogen bonding interactions also helps to stabilize the arrangement of α2 and α4. By virtue of being longer than Rap2b, Rap1b has an extension of five residues at the N-terminus and 13 at the C-terminus, the latter of which form a short helix α7. These segments of Rap1b are on the surface of the molecules at opposite ends of the dimer (Fig. 6).

The Rap proteins display an extensive dimerization interface (Fig. 6), consistent with the observation that each Rap protein exists as a stable dimer in solution (Fig. 4A and B). The Rap1b dimer uses almost 24% of the accessible surface area of a subunit in formation of the dimer. In the case of Rap2b the value is 20%. Such percentages are indicative of highly stable oligomers (Krissinel and Henrick, 2007). The dimer is stabilized by extensive van der Vaals interactions primarily involving aliphatic residues, in conjunction with hydrogen bonding interactions and solvent mediated bridging associations. The most important contributions are from side-chains on the concave surface, formed by α2, α3 and α5, interacting with the partner across the molecular twofold axis of symmetry. Additional interactions involve a self-association of the loop that links α2 and α3, together with the N-terminus α1 and α5. Helices α2 and α5 are thus not only critical for the fold of the Rap subunits but also to creating a suitable interface that leads to a highly stable dimer. This observation, together with the conserved disulphide bond (Fig. 7), ties in with the localization of Rap proteins to the oxidative environment of the periplasm where correct folding to support anti-toxin activity must occur.

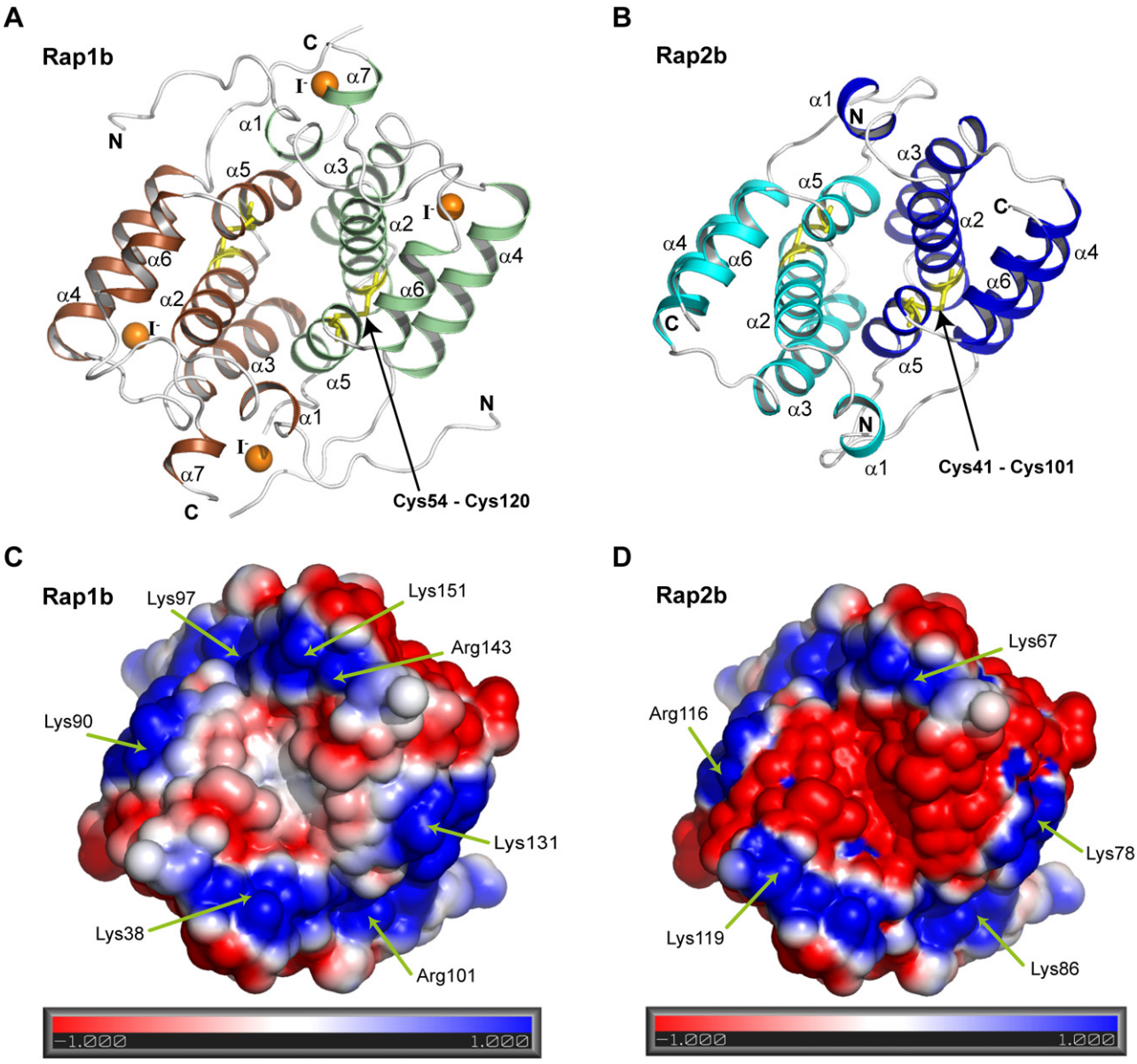


Fig. 6. Structural features of Rap1b and Rap2b dimers. The dimers are oriented to provide a view down their twofold axis of symmetry.
A. Ribbon diagram with helices of one Rap1b subunit coloured green, and the symmetry related molecule bronze. Elements of secondary structure, the N- and C-terminal residues are labelled and I^- ions are depicted as orange spheres. The disulphide bond is shown as yellow sticks and labelled.
B. Ribbon diagram of Rap2b with helices of one subunit blue, and the partner cyan.
C. and D. Electrostatic surface representation of Rap1b and Rap2b dimers in the same orientation as in (A) and (B). The electrostatic charge is contoured at 1 kT/e and −1 kT/e; negative (acidic) charge is red, positive (basic) blue. Residues that contribute to the basic patches are identified.

Importantly, sequence analysis and secondary structure predictions for the other Rap proteins, in *S. marcescens* and other organisms, suggest strongly that the novel fold revealed by our crystallographic analyses is generic for this entire family of proteins (Fig. 7). Not only are the main α-helices observed in both crystal structures closely mimicked by the structural predictions of the other members of the family, the Cys residues contributing to the structural disulphide bond are also conserved. Additionally, Rap2b shares slightly higher sequence identity with Rap2a (23% identity for the mature proteins) than it does with Rap1b (20% identity). Given the similarity of the Rap1b and 2b crystal structures, discussed above (Fig. S5), then this provides confidence that Rap2a shares this new fold and indeed that the Rap fold is a defining characteristic of this protein family, many

Table 1. Crystallographic statistics for the Rap1b and Rap2b structures.

	Rap1b	Rap2b
Space group	$P3_121$	$P222_1$
a, *b*, *c* (Å)	77.9, 77.9, 50.6	48.1, 57.0, 122.4
Resolution[a] (Å)	67.5–1.88 (1.98–1.88)	44.78–2.0 (2.1–2.0)
No. reflections recorded	162 739 (21 729)	271 849 (36 384)
Unique reflections	14 813 (2130)	23 666 (3390)
Completeness (%)	100.0	100.0
Multiplicity/$<I/\sigma I>$	11.0 (10.2)/26.6 (5.9)	11.5 (10.7)/25.7 (10.0)
Anomalous completeness (%)	100.0	100.0
Anomalous redundancy	5.6 (5.1)	6.0 (5.5)
Wilson *B* (Å^2)	28.3	20.6
Residues/waters/ligands	119/117/7	392/137/–
R_{merge}[b] (%)	5.3 (38.3)	8.5 (35.0)
R_{work}[c], R_{free}[d] (%)	18.3/22.1	18.9/24.8
Ave. *B*-factor (Å^2)		
Chain A, B, C, D	21.1	10.6, 9.8, 7.8, 8.9
Waters, iodides, ethylene glycol	42.3, 49.4, 55.7	14.6
Cruickshank DPI[e] (Å)	0.1	0.2
Ramachandran plot		
Most favoured	118 residues	381
Additional allowed	1	10
Outliers	0	Gln74
R.m.s.d. on ideal values		
Bond lengths (Å) angles (°)	0.02/1.51	0.02/1.79

a. Values in parentheses refer to the highest resolution shell.
b. $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - <I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $<I(hkl)>$ is the mean value of $I_i(hkl)$ for all *i* measurements.
c. $R_{work} = \Sigma_{hkl}||F_o| - |F_c||/\Sigma|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor.
d. R_{free} is the same as R_{cryst} except calculated with a subset, 5%, of data that are excluded from the refinement calculations.
e. Diffraction Precision Index (Cruickshank, 1999).
The PDB accession codes are 4AX2 (Rap1b) and 4B6I (Rap2b).

members of which are likely to represent T6SS immunity proteins.

Discussion

In this study we have identified two proteins, Ssp1 and Ssp2, as true (non-structural) secreted substrates of the *S. marcescens* T6SS and confirmed them as new T6SS-dependent antibacterial toxins. We have also demonstrated, genetically and biochemically, that the highly specific, cognate periplasmic immunity proteins, Rap1a and Rap2a, efficiently neutralize the effect of the Ssp proteins. The atomic structures of two other, related Rap proteins reveal that the immunity proteins should exhibit a novel protein fold likely only attained when they are present in the periplasm.

Our examination of Ssp1 and Ssp2 suggested that they represented novel antibacterial toxins, containing a domain of unknown function, DUF4285, present in prokaryotic and eukaryotic proteins. We were unable to detect significant sequence similarity with well-known peptidoglycan hydrolases and the structure prediction program Phyre2 (Kelley and Sternberg, 2009) was unable to assign them any peptidoglycan hydrolase-like structure. Nevertheless, as they are clearly periplasmic-acting toxins (Fig. 5) and non-resistant mutants showed apparent lytic and/or division defects (Figs 3A and S2), a cell wall targeting function seemed most likely. This idea has very recently been strongly supported by the observation that related proteins exhibit peptidoglycan amidase activity *in vitro* (see below; Russell *et al*., 2012). Importantly, through analysing secreted Ssp2 isolated from culture supernatant, we have also shown for the first time that T6-secreted effectors are not processed at either terminus during secretion. It seems clear that Ssp1 and Ssp2 are not the only effector proteins secreted by the *S. marcescens* T6SS. First, Ssp mutants are still able to kill *Pseudomonas fluorescens* as effectively as wild type Db10 (Fig. S4), implying that other toxins secreted in their absence are sufficient to maintain efficient antibacterial killing. Additionally, the susceptibility of the ΔT6SS mutant to self-targeting by the wild type strain is greater than that of the Δ*rap,ssp* mutant (Fig. 1A), implying additional immunity proteins and thus cognate secreted toxins within the T6SS gene cluster. Moreover, the magnitude of killing of the ΔT6SS mutant by wild type Db10 is less than that observed during T6-dependent killing of other organisms (Murdoch *et al*., 2011; Fig. S4), suggesting that additional secreted toxins (and cognate immunity proteins) are encoded elsewhere in the genome. Indeed, our unpub-

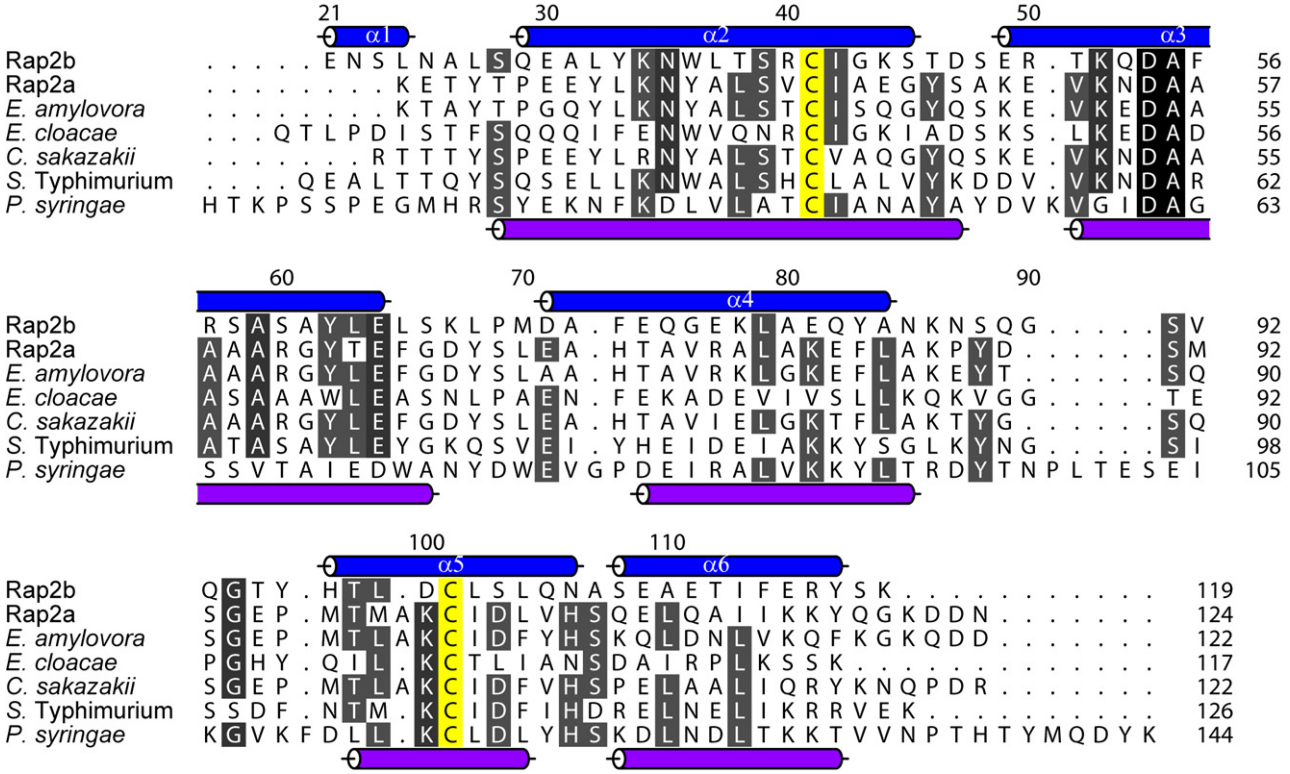


Fig. 7. Conservation of the Rap protein fold. Sequence alignment of *S. marcescens* Rap2b and Rap2a with homologous proteins from *Erwinia amylovora* (GenBank CBA22869.1), *Enterobacter cloacae* (NCBI Reference Sequence YP_003612051.1), *Cronobacter sakazakii* (NCBI Reference Sequence YP_001439955.1), *Salmonella* Typhimurium (NCBI Reference Sequence NP_459276.1) and *Pseudomonas syringae* (NCBI Reference Sequence YP_237109.1). The secondary structure of Rap2b (blue cylinders) and the predicted secondary structure of the *P. syringae* protein (purple cylinders) are shown. All proteins had similar predicted secondary structures (*PSIPRED*); the one shown is representative. Cysteine residues involved in disulphide bond formation are coloured yellow. Alignment was generated using *T-Coffee* and annotated using *ALINE*, using the mature proteins (i.e. without N-terminal signal peptides; numbering refers to the full-length proteins).

lished work has identified four other, unrelated, candidate substrates of the *Serratia* T6SS. Hence, we believe that *S. marcescens* uses a species-specific arsenal of secreted toxins to produce the potent and efficient targeting of a variety of competitor bacteria observed (Murdoch *et al.*, 2011).

Our work provides strong functional evidence for a new family of related T6SS substrates (Ssp1 and Ssp2 homologues) and family of related immunity proteins (Rap homologues) found in many different bacterial species and generally encoded within a main T6SS gene cluster of that organism. Ssp- and Rap-like proteins are found associated with a subset of T6SSs (some, but not all, closely related to the *S. marcescens* T6SS), but whether these T6SSs all exhibit antibacterial activity remains to be determined. While this report was in preparation, a bioinformatic study identified four disparate families of predicted T6-secreted peptidoglycan amidases, with Tse1 of *P. aeruginosa* being a member of 'Family 1' (Russell *et al.*, 2012). Entirely consistent with our data, Ssp-like proteins were recognized as one of these families, 'Family 4'. A related family of proposed cognate immunity proteins identified as co-occurring with all Ssp/Family 4 proteins is of course the Rap family proteins. Thus, in the nomenclature proposed by Russell *et al.*, Ssp1 and Ssp2 could be classified as Tae4.1^{SM} and Tae4.2^{SM}, and Rap1a and Rap2a as Tai4.1a^{SM} and Tai4.2a^{SM}. Unlike Families 1–3, the Family 4/Ssp proteins are almost unrecognizable as peptidoglycan amidases at a sequence and structure prediction level. Nevertheless, the purified Ssp homologue STM0277 from *Salmonella enterica* serovar Typhimurium was able to hydrolyse peptidoglycan cross-links at the D-Glu-*m*DAP bond of the acceptor stem (Russell *et al.*, 2012). Additionally, artificial expression and targeting of STM0277 to the periplasm was shown to be toxic to *E. coli*, with rescue by coexpression of the Rap homologue STM0278. However, studies were not conducted to show that STM0277 is a T6SS substrate or that it plays a role in T6-mediated antibacterial activity, neither was a role for STM0278 in self-resistance in the native, T6-elaborating organism investigated. In contrast, we have provided the comprehensive genetic, *in vivo* and biochemical data necessary to confirm that Ssp/Family 4 proteins are indeed a novel family of T6-secreted antibac-

terial toxins and moreover that the Rap family contains the cognate immunity proteins. Interestingly, it was recently reported that a mutant in the above Rap2a homologue in *S.* Typhimurium, STM0278, had a defect in replication in macrophages (Mulder *et al*., 2012). The reason behind this is not clear, but it may reflect the fact that this mutant has reduced fitness due to self-toxicity, just as we have shown for the Δ*rap2a* mutant. It should also be noted that Ssp–Rap pairs are not always found associated with T6SS genes. A particularly interesting example is the location of such a pair almost adjacent to genes encoding a classical RelE–RelB toxin–antitoxin pair (T-AT; Yamaguchi and Inouye, 2011) on a plasmid in *Acinetobacter baumannii* (Fig. 1). It is tempting to speculate that an original source of T6 toxin/resistance pairs is from plasmid T-AT systems.

The work of Russell *et al*. (2012) combined with our demonstration that Ssp1 and Ssp2 function as periplasmic-acting toxins is highly consistent with these proteins having a peptidoglycan amidase enzymatic activity. Of particular note, conserved Cys and His residues predicted by Russell *et al*. to represent the catalytic amino acids mediating peptidoglycan amide bond hydrolysis can be readily identified in Ssp1 and Ssp2. These are Cys50 (NTCAVRMS) and His133 (GHIDLIEP) in Ssp1, and Cys50 (NACAIRMS) and His131 in Ssp2 (GHATLWNG), with the equivalent, conserved regions in STM0277 being (NACPIRMS and GHVTLWNG). However, crucially, our data on the Ssp and Rap proteins in the context of T6SS-mediated attack and defence *in vivo* reveal that the situation is more subtle than this. In particular, Ssp1 and Ssp2 are clearly not redundant, despite having the same postulated enzymatic function. Rather, they have distinct activities or roles, with Ssp2 apparently more potent. In particular, we noted that the Ssp-dependent morphological phenotypes of the Δ*rap1a* and Δ*rap2a* mutants are different (Figs 3A and S2), that toxicity in the *E. coli* periplasm is only medium-dependent for Ssp1 (Fig. 5), and that the two may be relevant in different biological contexts (e.g. Ssp1 does not significantly contribute to self-targeting under the conditions of our standard assay, yet is clearly required for the self-toxicity observed in a Δ*rap1a* mutant). This specialization is consistent with a clear specificity for only the cognate Ssp–Rap partner, as we observed. In the native context, it is clear that none of the other three Rap proteins can confer cross-resistance to Ssp2 in the absence of Rap2a (nor did Ssp2 interact with Rap1a biochemically). The molecular basis for the difference between Ssp1 and Ssp2, which share 24% sequence identity, is not yet clear and will require further study, including determination of *in vitro* enzymatic activity and atomic structures.

The *S. marcescens* Rap proteins represent founder members of a new bacterial protein family, members of which represent immunity proteins for T6-secreted toxins, as exemplified by Rap1a and Rap2a. We have determined the structures of two members of this family, Rap1b and Rap2b, revealing a new protein fold. Given the sequence homology, conservation of key residues and shared predicted secondary structure throughout all members of the family (Fig. 7; Russell *et al*., 2012), this fold appears to be shared across members from different organisms. A conserved disulphide bond and the observed stable dimerization interface, together with canonical N-terminal signal sequences, are consistent with a periplasmic localization for all Rap family proteins. To date, the structures of two other T6 immunity proteins have been solved. The first is the cytoplasmic Tsi2 protein (Li *et al*., 2012; Zou *et al*., 2012). Like the Rap proteins, Tsi2 exhibits a helical fold, is acidic and exists as a stable dimer in solution. However, the structures of Tsi2 and the Rap proteins are unrelated. Second, and very recent, is the structure of the periplasmic Tsi1 protein, complexed with the secreted peptidoglycan hydrolase effector, Tse1 (Ding *et al*., 2012). Tse1 possesses a strikingly accessible active site, facilitating its promiscuous and toxic peptidoglycan amidase activity (Chou *et al*., 2012). Tsi1 binds to Tse1 in a 1:1 complex, occluding the substrate-binding site of Tse1 in order to neutralize its activity (Ding *et al*., 2012). Critically, although the Rap1a/Rap2a proteins and Tsi1 both mediate resistance to peptidoglycan hydrolase toxins, their structures now appear to be entirely unrelated. Tsi1 exhibits an all β fold, related to a classical β-propeller, whereas Rap family proteins exhibit a novel helical fold. This indicates that the mechanisms by which immunity proteins confer resistance may be divergent even among those with effectors of similar function. Rap1a and Rap2a have an obvious immunity phenotype specific to their cognate secreted toxins, Ssp1 and Ssp2. However, the role of Rap1b and Rap2b is not yet known. While they do not appear to play a role in self-resistance, they may play a role in resistance towards closely related bacteria secreting similar toxins. It is worth noting that other organisms also possess multiple Rap family proteins for a given Ssp family protein (Fig. 1C and data not shown); therefore, whatever the function of Rap1b and 2b turns out to be, it may not be unique to *Serratia*.

Complementing our genetic and phenotypic demonstration of the cognate toxin-immunity function of Ssp1–Rap1a and Ssp2–Rap2a, we report detailed biochemical characterization of the interactions between these purified T6-secreted toxins and immunity proteins. For efficient self-protection, these interactions should be tight and highly specific, as was observed. Formation of Ssp1–Rap1a and Ssp2–Rap2a complexes is exothermic, with low nanomolar K_d, and with a stoichiometry of 2:2. A binding affinity of this order agrees well with the K_d of 3 nM reported for the Tsi1–Tse1 interaction (Ding *et al*., 2012);

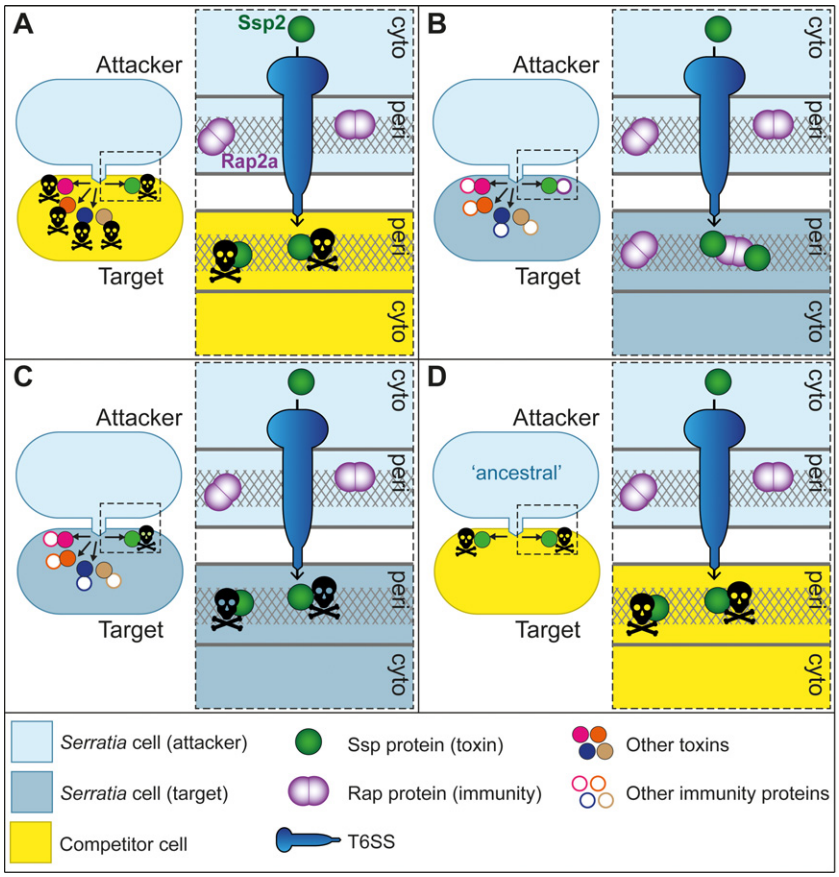


Fig. 8. Model for action and context of Ssp and Rap proteins.
A. Wild type *S. marcescens* Db10 uses its T6SS to inject multiple different antibacterial toxins (solid circles), including Ssp2 (green), into a susceptible target competitor cell, cumulatively causing a rapid death. Ssp2 attacks the cell wall in the periplasm (peri); other toxins are likely to attack targets in the cytoplasm (cyto).
B. Another wild type cell is resistant to T6SS attack by its neighbour because of the presence of cognate immunity proteins for all the toxins (open circles). Rap2a dimers (purple) form a protective barrier in the periplasm, rapidly binding and sequestering Ssp2.
C. When a Δ*rap2a* mutant of Db10 is the target strain, all of the effectors except for Ssp2 are still neutralized, but Ssp2 secreted by wild type cells is now able to cause toxicity in the target.
D. In a putative ancestral cell, T6SS secretion of Ssp2, but not other toxins later acquired by horizontal transfer, inhibits a competitor target cell. In this case, loss of Ssp2 (or resistance to Ssp2 in the target) would prevent inhibition. For clarity, Ssp1 and Rap1a are not shown but would behave similarly to Ssp2 and Rap2a.

however, the Tsi1–Tse1 complex has a stoichiometry of 1:1, again highlighting significant differences between different pairs of toxin-immunity proteins. The structure of a Ssp–Rap complex and arrangement of the subunits has yet to be determined, although it is likely that the two Ssp proteins bind to the same part of each Rap monomer, exploiting the twofold symmetry. Li *et al*. suggest that Tse2 interacts with an acidic patch on Tsi2 distal to the dimer interface (Li *et al*., 2012). They also note that, like many toxin–antitoxin pairs, the resistance proteins Tsi1–3 are more acidic than the toxin proteins Tse1–3. This pattern is strikingly followed with the Ssp (theoretical pI 9) and Rap (theoretical pI 5–6) proteins. Nevertheless, while the highly acidic Rap proteins bind their cognate Ssp partners, which are noticeably basic, with high affinity, the story is not as simple as charge complementarity. Surprisingly for proteins with such low pI values, the striking surface feature conserved in Rap1b and Rap2b structures is a crescent-shaped basic patch on either side of the dimer (Fig. 6C and D). It is possible that such a basic crescent, if present on toxin-binding Rap proteins, may contribute to orientation of the partner during binding; conversely, it might contribute to the lack of Ssp binding by Rap1b and Rap2b. Additionally, all four *S. marcescens* Rap proteins are acidic and both Ssp proteins basic, yet only two, highly specific interactions are observed (Rap1a–Ssp1 and Rap2a–Ssp2). While charge complementarity would be expected to be important for specific complex formation, it is likely that shape considerations play an equally important role. The structure of a toxin-immunity protein complex will be necessary to delineate the molecular features that govern specific Rap–Ssp association. We are working towards that goal.

In summary, our data support a model (Fig. 8) whereby Ssp1 and Ssp2 are toxins secreted by the *S. marcescens* T6SS into the periplasm of a neighbouring cell. If this is an isogenic sibling, the cognate Rap proteins provide an efficient protective barrier, effectively binding and sequestering the toxin, whereas if it is a competitor, Ssp1/2 are free to attack the cell wall. However, Ssp1 and Ssp2 are only two of multiple distinct T6-secreted toxins. This provides the observed redundancy of function: if Ssp1/2 are missing, the other toxins still cause the efficient death of a competitor. This 'belt and braces' approach provides great robustness: if a competitor becomes resistant to one or two toxins, the others will still provide the ability to kill or inhibit the competitor efficiently. Of course, the toxic effect of Ssp1/2 can be seen against self (if the cognate Rap is missing) as Db10 has resistance proteins to neutralize the other toxins. We believe that other effectors most likely

act on different cellular targets, analogous to the *Pseudomonas*-specific, cytoplasmic-acting toxin, Tse2 (Li *et al*., 2012). Finally, we speculate that as the Ssp and Rap proteins are encoded within the T6SS gene cluster, and related genes are found associated with other T6SSs, particularly closely related ones, they may represent the ancestral substrates of the *S. marcescens* T6SS. Subsequently, other toxins could have been horizontally acquired from diverse sources and adopted by this highly versatile secretion machine. The exciting task of deciphering how such species- and strain-specific arsenals of toxins and immunity proteins are co-ordinated by the producing cell and how they contribute to the dynamic composition of polymicrobial communities in infection and the environment lies ahead.

Experimental procedures

Bacterial strains, plasmids and culture conditions

All strains and plasmids used in this study are detailed in Table S1. All mutants constructed in *S. marcescens* Db10 were in-frame deletion mutants, generated by allelic exchange as described previously (Murdoch *et al*., 2011). Streptomycin-resistant derivatives were generated by phage ϕIF3-mediated transduction of the resistance allele from *S. marcescens* Db11 (Petty *et al*., 2006). Plasmids for constitutive expression of proteins in *S. marcescens* were derived from pSUPROM, plasmids for arabinose-inducible protein expression were derived from pBAD18-Kn, and derivatives of the pET15b-TEV plasmid were generated for protein overexpression and purification. *S. marcescens* was grown at 30°C in LB (10 g l^{-1} tryptone, 5 g l^{-1} yeast extract, 10 g l^{-1} NaCl, with 1.5 g l^{-1} agar for solid media) or minimal media (40 mM K_2HPO_4, 15 mM KH_2PO_4, 0.1% $(NH_4)_2SO_4$, 0.4 mM $MgSO_4$, 0.2% glucose) and *E. coli* was normally grown at 37°C in LB or M9 minimal media (M9). M9 (Sambrook and Russell, 2001) contained 0.5% glycerol (plus stated concentrations of arabinose and glucose as required). Growth analyses were performed in 96-well plates in a BioTek Synergy platereader. When required, media were supplemented with antibiotics: ampicillin (Ap) 100 μg ml^{-1}, kanamycin (Kn) 100 μg ml^{-1}, streptomycin (Sm) 100 μg ml^{-1}, chloramphenicol (Cm) 25 μg ml^{-1}; to maintain repression of proteins expressed from pBAD18-Kn, 0.5% glucose was added to the media for cloning and maintenance.

Antibacterial competition/co-culture assays

These were based on the assay described previously (Murdoch *et al*., 2011). In brief, the attacker strain and target strain (both at OD_{600} 0.5) were mixed at an initial ratio of 1:1 attacker : target, co-cultured on solid LB for 7.5 h at 30°C and then the surviving target cells enumerated by serial dilution and viable counts on streptomycin-containing media. The target strain was always the streptomycin-resistant version of the mutant in question (Table S1). Statistical significance testing was performed by ANOVA followed by Dunnett's post test (GraphPad Prism software).

Bioinformatic identification of homologues of Rap and Ssp proteins and determination of their genetic contexts utilized the BLAST servers and sequence databases at the NCBI (http://www.ncbi.nlm.nih.gov).

Immunodetection of secreted proteins

Anti-Hcp immunoblots were performed as described (Murdoch *et al*., 2011). For detection of Ssp1 and Ssp2, cellular and secreted fractions were prepared from 25 ml of culture grown for 7 h in LB. Secreted proteins were precipitated using 50:50 chloroform : methanol followed by methanol wash and resuspension in 2× gel sample buffer (Murdoch *et al*., 2011). Cellular samples were prepared by sonicating harvested cells in 20 mM Tris·HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100 and isolating the soluble fraction by centrifugation. Anti-Ssp1 and anti-Ssp2 rabbit polyclonal antibodies were raised to the purified proteins (Eurogentec, Belgium) and used at 1:1000; peroxidase-conjugated secondary (Thermo Fisher Scientific) was used at 1:10 000. Anti-RNAP β (Neoclone, USA) was used at 1:20 000, with anti-mouse secondary (Roche) at 1:10 000. In all cases, protein from the same number of cells was loaded in the secreted versus the cellular samples.

Localization of Rap*-HA proteins*

Fractionation was performed using a cold osmotic shock procedure. Following growth of cultures for 5 h in LB, Tris·HCl pH 7.8 was added to 5 ml of cells to a final concentration of 50 mM and the cells were incubated for 10 min at room temperature, then recovered by centrifugation and washed once in LB. Cell pellets were resuspended in 1 ml of 40% sucrose, 30 mM Tris·HCl pH 7.8, 2 mM EDTA and incubated for 10 min at 30°C. One hundred microlitres of this fraction ('whole cell') was removed for analysis. Remaining cells were recovered by centrifugation, resuspended in 900 μl of ice-cold water and incubated on ice for 10 min, resulting in the release of the periplasm. After centrifugation, 100 μl of the supernatant ('periplasm' fraction) was retained for analysis. The pellet, containing the 'cytoplasm + membranes' fraction, was resuspended in 900 μl of 50 mM Tris·HCl pH 7.8 and 100 μl retained for analysis. Equivalent amounts, on a per cell basis, of each fraction in each strain were assayed. Anti-RNAP β was used as above, anti-MBP (NEB) was used at 1:10 000 and anti-HA (Roche) was used at 1:6000, all with anti-mouse secondary as above.

Microscopic analysis of colony and cell morphology

Overnight cultures were normalized to OD_{600} 0.5, diluted 10^{-2} (culture spots) or to obtain single colonies, 10 μl spotted onto solid media and grown for 24 or 48 h. Macroscopic morphology of culture spots and single colonies were recorded using a Zeiss MZ16FA Stereo Microscope with a Leica DFC350 FX camera and Leica AF6000 software. Microscopic analysis of single cells taken from spots grown on solid media as above was performed by Differential Interference Contrast (DIC) microscopy using an Axioskop 2 mot plus (Zeiss) with a SPOT RT KE camera and SPOT software (Diagnostic Instruments).

Protein purification and in vitro *analysis*

A full description of protein purification is provided in the *Supporting information*. In brief, proteins were overproduced in *E. coli* with a TEV protease-cleavable His_6 tag and isolated by Ni^{2+} affinity purification. If required, the His_6 tag was cleaved using TEV protease and the protein re-isolated by reverse Ni^{2+} chromatography. A final SEC step was then always performed. For SEC analysis of complex formation, His_6-tagged proteins (10 nmol each) in 50 mM Tris·HCl, pH 7.5, 250 mM NaCl were separated on a Superdex 75 10/300 GL column, calibrated using molecular weight standards (GE Healthcare). Quantitative SYPRO Orange staining was performed as described (Rickman *et al*., 2004), with image analysis using ImageJ. Molar ratios were the mean of at least four quantifications. ITC was performed in 50 mM Tris·HCl, pH 7.5, 250 mM NaCl at 30°C in a MicroCal iTC200 calorimeter. The sample cell contained 6 μM Rap1a dimer or Rap2a dimer and the syringe contained 120 μM Ssp1 or Ssp2. Titrations consisted of 30 × 8 μl injections of Ssp into Rap (or into buffer alone as control). Data analysis was performed with the Origin software provided (MicroCal). For affinity isolation of Ssp1 and Ssp2, 10 μg of His_6-tagged Rap protein was immobilized on magnetic Ni^{2+} beads (Qiagen), incubated for 1 h with culture supernatant (after 7 h growth in LB), beads washed and bound proteins eluted by the addition of gel sample buffer. Identified proteins were excised from the Coomassie- (Ssp2) or Silver- (Ssp1) stained gel and identified by mass spectrometry.

Crystallographic analyses

Well-ordered trigonal and orthorhombic crystals of Rap1b and Rap2b, respectively, were obtained. The asymmetric unit for Rap1b consisted of a single subunit, while Rap2b displayed two dimers in the asymmetric unit. Diffraction data were recorded in-house and experimental phases were derived by single-wavelength anomalous diffraction measurements (Micossi *et al*., 2002), exploiting the signal from endogenous S atoms and I^- ions that had been added by soaking. The electron density maps were of high quality, and the models were completed, then refined to high resolution using standard methods (Dawson *et al*., 2008). Full crystallographic details are provided in the *Supporting information* and Table 1. Atomic co-ordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession codes 4AX2 (Rap1b) and 4B6I (Rap2b).

Acknowledgements

This work was supported by the Royal Society of Edinburgh (Personal Research Fellowship to S. J. C.), Medical Research Scotland (382 FRG), the Wellcome Trust [grants 082596, 083481, a PhD studentship (G. E.) and an ISSF award to the University of Dundee], the European Commission (FP7/2007-2013, the Aeropath project) and a Royal Society Project Grant (S. J. C.). We would like to thank Maximilian Fritsch, Sarah Murdoch, Matthias Trost, Emma Compton, Arnaud Javelle, Frank Sargent and members of the WNH lab for technical assistance and discussions.

References

Aubert, D.F., Flannagan, R.S., and Valvano, M.A. (2008) A novel sensor kinase-response regulator hybrid controls biofilm formation and type VI secretion system activity in *Burkholderia cenocepacia*. *Infect Immun* **76:** 1979–1991.

Basler, M., Pilhofer, M., Henderson, G.P., Jensen, G.J., and Mekalanos, J.J. (2012) Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483:** 182–186.

Bonemann, G., Pietrosiuk, A., and Mogk, A. (2010) Tubules and donuts: a type VI secretion story. *Mol Microbiol* **76:** 815–821.

Burtnick, M.N., Brett, P.J., Harding, S.V., Ngugi, S.A., Ribot, W.J., Chantratita, N., *et al*. (2011) The cluster 1 type VI secretion system is a major virulence determinant in *Burkholderia pseudomallei*. *Infect Immun* **79:** 1512–1525.

Cascales, E. (2008) The type VI secretion toolkit. *EMBO Rep* **9:** 735–741.

Cascales, E., and Cambillau, C. (2012) Structural biology of type VI secretion systems. *Philos Trans R Soc Lond B Biol Sci* **367:** 1102–1111.

Choi, S.H., Lee, J.E., Park, S.J., Kim, M.N., Choo, E.J., Kwak, Y.G., *et al*. (2007) Prevalence, microbiology, and clinical characteristics of extended-spectrum beta-lactamase-producing *Enterobacter* spp., *Serratia marcescens*, *Citrobacter freundii*, and *Morganella morganii* in Korea. *Eur J Clin Microbiol Infect Dis* **26:** 557–561.

Chou, S., Bui, N.K., Russell, A.B., Lexa, K.W., Gardiner, T.E., Leroux, M., Vollmer, W., and Mougous, J.D. (2012) Structure of a peptidoglycan amidase effector targeted to Gram-negative bacteria by the Type VI secretion system. *Cell Rep* **1:** 656–664.

Cruickshank, D.W. (1999) Remarks about protein structure precision. *Acta Crystallogr D Biol Crystallogr* **55:** 583–601.

Dawson, A., Fyfe, P.K., and Hunter, W.N. (2008) Specificity and reactivity in menaquinone biosynthesis: the structure of *Escherichia coli* MenD (2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexadiene-1-carboxylate synthase). *J Mol Biol* **384:** 1353–1368.

Ding, J., Wang, W., Feng, H., Zhang, Y., and Wang, D.C. (2012) Structural insights into the *Pseudomonas aeruginosa* type VI virulence effector Tse1 bacteriolysis and self-protection mechanisms. *J Biol Chem* **287:** 26911–26920.

Filloux, A., Hachani, A., and Bleves, S. (2008) The bacterial type VI secretion machine: yet another player for protein transport across membranes. *Microbiology* **154:** 1570–1583.

Gerlach, R.G., and Hensel, M. (2007) Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens. *Int J Med Microbiol* **297:** 401–415.

Hood, R.D., Singh, P., Hsu, F., Guvener, T., Carl, M.A., Trinidad, R.R., *et al*. (2010) A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* **7:** 25–37.

Jani, A.J., and Cotter, P.A. (2010) Type VI secretion: not just for pathogenesis anymore. *Cell Host Microbe* **8:** 2–6.

Kelley, L.A., and Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4:** 363–371.

Krissinel, E., and Henrick, K. (2007) Inference of macromo-

lecular assemblies from crystalline state. *J Mol Biol* **372:** 774–797.
Laskowski, R.A., Watson, J.D., and Thornton, J.M. (2005) ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res* **33:** W89–W93.
Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzki, S., *et al*. (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci USA* **106:** 4154–4159.
Li, M., Le Trong, I., Carl, M.A., Larson, E.T., Chou, S., De Leon, J.A., *et al*. (2012) Structural basis for type VI secretion effector recognition by a cognate immunity protein. *PLoS Pathog* **8:** e1002613.
Lockhart, S.R., Abramson, M.A., Beekmann, S.E., Gallagher, G., Riedel, S., Diekema, D.J., *et al*. (2007) Antimicrobial resistance among Gram-negative bacilli causing infections in intensive care unit patients in the United States between 1993 and 2004. *J Clin Microbiol* **45:** 3352–3359.
MacIntyre, D.L., Miyata, S.T., Kitaoka, M., and Pukatzki, S. (2010) The *Vibrio cholerae* type VI secretion system displays antimicrobial properties. *Proc Natl Acad Sci USA* **107:** 19520–19524.
Micossi, E., Hunter, W.N., and Leonard, G.A. (2002) De novo phasing of two crystal forms of tryparedoxin II using the anomalous scattering from S atoms: a combination of small signal and medium resolution reveals this to be a general tool for solving protein crystal structures. *Acta Crystallogr D Biol Crystallogr* **58:** 21–28.
Mulder, D.T., Cooper, C.A., and Coombes, B.K. (2012) Type VI secretion system-associated gene clusters contribute to pathogenesis of *Salmonella enterica* serovar Typhimurium. *Infect Immun* **80:** 1996–2007.
Murdoch, S.L., Trunk, K., English, G., Fritsch, M.J., Pourkarimi, E., and Coulthurst, S.J. (2011) The opportunistic pathogen *Serratia marcescens* utilizes type VI secretion to target bacterial competitors. *J Bacteriol* **193:** 6057–6069.
de Pace, F., Nakazato, G., Pacheco, A., de Paiva, J.B., Sperandio, V., and da Silveira, W.D. (2010) The type VI secretion system plays a role in type 1 fimbria expression and pathogenesis of an avian pathogenic *Escherichia coli* strain. *Infect Immun* **78:** 4990–4998.
Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8:** 785–786.
Petty, N.K., Foulds, I.J., Pradel, E., Ewbank, J.J., and Salmond, G.P. (2006) A generalized transducing phage (phiIF3) for the genomically sequenced *Serratia marcescens* strain Db11: a tool for functional genomics of an opportunistic human pathogen. *Microbiology* **152:** 1701–1708.
Pukatzki, S., Ma, A.T., Revel, A.T., Sturtevant, D., and Mekalanos, J.J. (2007) Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc Natl Acad Sci USA* **104:** 15508–15513.
Pukatzki, S., McAuley, S.B., and Miyata, S.T. (2009) The type VI secretion system: translocation of effectors and effector-domains. *Curr Opin Microbiol* **12:** 11–17.
Rickman, C., Meunier, F.A., Binz, T., and Davletov, B. (2004) High affinity interaction of syntaxin and SNAP-25 on the plasma membrane is abolished by botulinum toxin E. *J Biol Chem* **279:** 644–651.
Rosales-Reyes, R., Skeldon, A.M., Aubert, D.F., and Valvano, M.A. (2012) The Type VI secretion system of *Burkholderia cenocepacia* affects multiple Rho family GTPases disrupting the actin cytoskeleton and the assembly of NADPH oxidase complex in macrophages. *Cell Microbiol* **14:** 255–273.
Russell, A.B., Hood, R.D., Bui, N.K., LeRoux, M., Vollmer, W., and Mougous, J.D. (2011) Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **475:** 343–347.
Russell, A.B., Singh, P., Brittnacher, M., Bui, N.K., Hood, R.D., Carl, M.A., *et al*. (2012) A widespread bacterial type VI secretion effector superfamily identified using a heuristic approach. *Cell Host Microbe* **11:** 538–549.
Sambrook, J., and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
Schwarz, S., West, T.E., Boyer, F., Chiang, W.C., Carl, M.A., Hood, R.D., *et al*. (2010) *Burkholderia* type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *PLoS Pathog* **6:** e1001068.
Suarez, G., Sierra, J.C., Erova, T.E., Sha, J., Horneman, A.J., and Chopra, A.K. (2010) A type VI secretion system effector protein, VgrG1, from *Aeromonas hydrophila* that induces host cell toxicity by ADP ribosylation of actin. *J Bacteriol* **192:** 155–168.
Wiseman, T., Williston, S., Brandts, J.F., and Lin, L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **179:** 131–137.
Yamaguchi, Y., and Inouye, M. (2011) Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat Rev Microbiol* **9:** 779–790.
Zheng, J., and Leung, K.Y. (2007) Dissection of a type VI secretion system in *Edwardsiella tarda*. *Mol Microbiol* **66:** 1192–1206.
Zou, T., Yao, X., Qin, B., Zhang, M., Cai, L., Shang, W., *et al*. (2012) Crystal structure of *Pseudomonas aeruginosa* Tsi2 reveals a stably folded superhelical antitoxin. *J Mol Biol* **417:** 351–361.

Supporting information

Additional supporting information may be found in the online version of this article.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.